

Sun Sep 7, 09:25:59 2003

us-10-056-884-1.rge

Page 1.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 5, 2003, 19:09:56 / Search time 12228 Seconds  
(without alignments)

11602.436 Million cell updates/sec

Title: US-10-056-884-1

Sequence: 1 caagcactgtctgaagtgct.....aaaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Minimum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

GenEmbl:  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.pro:\*  
11: gb.srs:\*  
12: gb.sr:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.or:\*  
21: em.ov:\*  
22: em.pat:\*  
23: em.ph:\*  
24: em.pl:\*  
25: em.ro:\*  
26: em.srs:\*  
27: em.un:\*  
28: em.vi:\*  
29: em.yi:\*  
30: em.htg.hum:\*  
31: em.htg.in:\*  
32: em.htg.other:\*  
33: em.htg.mus:\*  
34: em.htg.pln:\*  
35: em.htg.fed:\*  
36: em.htg.mam:\*  
37: em.htg.vrt:\*  
38: em.htg.hum:\*  
39: em.htgo.hum:\*  
40: em.htgo.mus:\*  
41: em.htgo.other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3198.4	92.2	5646	9	AB037738 Homo sapi
2	2053.4	59.0	182638	2	AC019335 Homo sapi
3	2053.4	58.9	184589	9	AC008716 Homo sapi
4	1640.8	58.8	98360	9	AC008473 Homo sapi
5	1640.8	47.3	171942	6	AX405760 Sequence
6	1161.8	33.5	209114	9	AC008652 Homo sapi
7	1161.8	33.5	209114	9	AC008383 Homo sapi
8	853.2	24.6	135132	2	AC117282
9	853.2	24.6	135132	2	AC117284
10	819.4	23.6	242879	2	AC117827
11	811.8	23.4	230128	10	AC098707
12	806.2	23.2	249703	2	AC112539
13	367.4	10.6	781	10	BC049734 Mus muscu
14	319.2	9.2	175059	2	BC432465
15	317.6	9.2	200467	2	EX470157
16	317.6	8.3	230261	2	EX530085
17	286.4	8.3	184319	2	EX511303
18	286.4	8.2	192400	2	EX530406
19	278.4	8.0	83028	2	EX004755
20	278.4	8.0	243835	5	AL935304
21	255.4	7.4	174712	10	AL831725
22	251.6	7.3	1890	5	AY120891
23	220.8	6.4	183038	2	AC107770
24	216.4	6.2	246164	2	AC098751
25	216.4	6.2	322972	2	AC129853
26	216.4	6.1	138872	2	AC113951
27	189.6	5.5	243299	2	AC113800
28	188	5.4	3086	9	BD109391
29	167	4.8	109201	9	BC013764
30	167	4.8	169362	2	AC000403
31	167	4.8	251187	2	AL136440
32	165.4	4.8	145979	2	AL358875
33	159.8	4.6	145979	2	AC102815
34	156.6	4.5	217336	2	AC113144
35	149.8	4.3	675	10	BC049679
36	129.2	3.7	256933	10	AC099771
37	128.2	3.7	194240	2	AC118142
38	127.6	3.7	203720	9	AC093861
39	114.8	3.3	569	5	AY093634
40	104.6	3.0	100029	10	AB014174_3
41	84.4	2.4	2093	5	BC051776
42	81.8	2.4	1450	17	AF132205
43	80.8	2.4	2689	9	BC032473
44	79.2	2.3	246548	2	EX530062
45	79.2	2.3	4582	9	BC037864

## ALIGNMENTS

RESULT 1	AB037738	5646 bp	linear	PRI 14-MAR-2000
LOCUS	AB037738			
DEFINITION	Homo sapiens mRNA for KIAA1317 protein, partial cds.			
ACCESSION	AB037738			
VERSION	AB037738.1	GI:7243014		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
AUTHORS	1 (sates) Kikuno, R., Ishikawa, K.I., Hirotsawa, M. and Ohara, O.			
TITLE	Nagase, T., Kikuno, R., Ishikawa, K.I., Hirotsawa, M. and Ohara, O. Prediction of the coding sequences of newly identified human genes.			



Db	2260	CAATGAGAGAGAGCTGGAGAAATGTATCCAGATTTCTTAAAAATCAAAATTCAGATC	2319
Qy	1743	GGTTCTGAGAGAAAAATCTTGCAATCTGAATCTTTTAAAGAGATCATCTATAAG	1802
Db	2320	GGTTCTGAGAGAAAAATCTTGCAATCTGAATCTTTTAAAGAGATCATCTATAAG	2379
Qy	1803	GGAGGGCTGGGGCGGGGAAAAAAGAGCATTTTGAATTAACCTCATAAA	1862
Db	2380	GGAGGGCTGGGGCGGGGAAAAAAGAGCATTTTGAATTAACCTCATAAA	2438
Qy	1863	GAATTCATATTTTAAAGAAAAAATACATATATGATGACATTTCTTGAACAAT	1922
Db	2439	GAATTCATATTTTAAAGAAAAAATACATATATGATGACATTTCTTGAACAAT	2498
Qy	1923	AGTCATATATATCTAGTGCCTACTTACCTTACCTTAAATGATTAATCCAGAG	1982
Db	2499	AGTCATATATATCTAGTGCCTACTTACCTTACCTTAAATGATTAATCCAGAG	2558
Qy	1983	GTAGATTTCTTTCTAGATGTGGAAGTACAGAAAAATCTTTTATGTTATTTGTTTA	2042
Db	2559	GTAGATTTCTTTCTAGATGTGGAAGTACAGAAAAATCTTTTATGTTATTTGTTTA	2618
Qy	2043	CTTGCTCCATGTGCTAACTATCTTATATATATAGAGCCAGCTAGTAAAGTAGCTG	2102
Db	2619	CTTGCTCCATGTGCTAACTATCTTATATATATAGAGCCAGCTAGTAAAGTAGCTG	2678
Qy	2103	AGAGGCTTGGAGAGCATTTATCCCAAACTGGGTTTCTCTCATCTTACCTCCCT	2162
Db	2679	AGAGGCTTGGAGAGCATTTATCCCAAACTGGGTTTCTCTCATCTTACCTCCCT	2738
Qy	2163	CTTTGAAATGAGGGATATGTAAGAAAAAGATCTGGCCCAATGACATTAAGTTTGA	2222
Db	2739	CTTTGAAATGAGGGATATGTAAGAAAAAGATCTGGCCCAATGACATTAAGTTTGA	2798
Qy	2223	AAATTTGGTTTCTCTTTGTTTATGAGGTTGGAGGAAATGACATTTATGACTTTT	2282
Db	2799	AAATTTGGTTTCTCTTTGTTTATGAGGTTGGAGGAAATGACATTTATGACTTTT	2858
Qy	2283	CACTCAAATCTATATGTCCAGTTTATATGATCCGATGATGATGATTTGTGACAA	2342
Db	2859	CACTCAAATCTATATGTCCAGTTTATATGATCCGATGATGATGATTTGTGACAA	2918
Qy	2343	CAAGCAAACTATATGTATATATACATGACGACAGAGGCCAGGAGCTTACCTCC	2402
Db	2919	CAAGCAAACTATATGTATATATACATGACGACAGAGGCCAGGAGCTTACCTCC	2978
Qy	2403	AAAGGCTGCTCCGCTCCGAGAGCCCTCTTGAATATTTCAATGATGATGACTTC	2462
Db	2979	AAAGGCTGCTCCGCTCCGAGAGCCCTCTTGAATATTTCAATGATGATGACTTC	3038
Qy	2463	TGACTCTTTCTTAAATTTCTTTGGAGATTTCCAGGCTTTCTTCAACAATTTCTA	2522
Db	3039	TGACTCTTTCTTAAATTTCTTTGGAGATTTCCAGGCTTTCTTCAACAATTTCTA	3098
Qy	2523	ACATCAAAATGACTCTCATCAACAATTTGTAATCTTATGTAATTAATACCTCA	2582
Db	3099	ACATCAAAATGACTCTCATCAACAATTTGTAATCTTATGTAATTAATACCTCA	3158
Qy	2583	GGCTCCATTTTACTGCTTGTGCTTTGTGCTTAAAGAGAGATGAGAGAGCTGTCA	2642
Db	3159	GGCTCCATTTTACTGCTTGTGCTTTGTGCTTAAAGAGAGATGAGAGAGCTGTCA	3218
Qy	2643	AACATTCCTTGTGTAAATAAATCAAAATCTTATATCAAAAATTTTCTGCTAAATGAC	2702
Db	3219	AACATTCCTTGTGTAAATAAATCAAAATCTTATATCAAAAATTTTCTGCTAAATGAC	3278
Qy	2703	TTCGACACTGAGGCTCTGTAACCGTGAACGTAATTAACCCCTTTTCCATGTTTCAAA	2762
Db	3279	TTCGACACTGAGGCTCTGTAACCGTGAACGTAATTAACCCCTTTTCCATGTTTCAAA	3338
Qy	2763	GTTCCTTAGCGCCAGATTAAATGATGCTGCTTTCACATATATCCATTAATGTTCTGT	2822
Db	3339	GTTCCTTAGCGCCAGATTAAATGATGCTGCTTTCACATATATCCATTAATGTTCTGT	3398
Qy	2823	GTTCCTTAGCTGTGTGGAACCTAAGGACAAACACACAGTACTTGAATTAAGGTCGGGCTT	2882
Db	3399	GTTCCTTAGCTGTGTGGAACCTAAGGACAAACACACAGTACTTGAATTAAGGTCGGGCTT	3458
Qy	2883	TTCCTTGTGTTTGAAGAAAGTGTATTCACACACAACTAATTAATTTCTAATAAATTT	2942
Db	3459	TTCCTTGTGTTTGAAGAAAGTGTATTCACACACAACTAATTAATTTCTAATAAATTT	3518
Qy	2943	TAAACTAACAGCTACATTTTACTTGTGAGCCGTTTGTGTTGCTTGGGATTCG	3002
Db	3519	TAAACTAACAGCTACATTTTACTTGTGAGCCGTTTGTGTTGCTTGGGATTCG	3078
Qy	3003	GGCTTGGCTGTGCTCCATGCTAGGATTTAGCTGTGCTATTTTATGATGCTGTAAACAC	3062
Db	3579	GGCTTGGCTGTGCTCCATGCTAGGATTTAGCTGTGCTATTTTATGATGCTGTAAACAC	3638
Qy	3063	CCAAACAGGTAACAGTCAAGCTCCAGAGTTAAGGTTTCAATTTCTAATGAACTATCTT	3122
Db	3639	CCAAACAGGTAACAGTCAAGCTCCAGAGTTAAGGTTTCAATTTCTAATGAACTATCTT	3698
Qy	3123	TTCATTTACATCTGACTGTGATATGACACAGCCAAAGAACTGTTAATAGCCATCCGT	3182
Db	3699	TTCATTTACATCTGACTGTGATATGACACAGCCAAAGAACTGTTAATAGCCATCCGT	3758
Qy	3183	CCATGTAACTGCTGATTTTCTAAGGTACCAATACCTTTTCAATGATGATGATGATGAT	3242
Db	3759	CCATGTAACTGCTGATTTTCTAAGGTACCAATACCTTTTCAATGATGATGATGATGAT	3818
Qy	3243	AAAGTTAAAGACAGCTTTTATTTTTCAGCATCTCTCATGATCTTCAAGTGTAAACAAA	3302
Db	3819	AAAGTTAAAGACAGCTTTTATTTTTCAGCATCTCTCATGATCTTCAAGTGTAAACAAA	3878
Qy	3303	ATTAATTTGCTAATTAATAGTGTGTGTCACAAAGCACTCTTAATTTTATTTGCTGTGTG	3362
Db	3879	ATTAATTTGCTAATTAATAGTGTGTGTGTCACAAAGCACTCTTAATTTTATTTGCTGTGTG	3938
Qy	3363	TGCATGTGTATGTGTATCATAGTAAATTAAGCAATGATGATTAATAAATAAATAA	3422
Db	3939	TGCATGTGTATGTGTATCATAGTAAATTAAGCAATGATGATTAATAAATAAATAA	3998
Qy	3423	AAAAAAAAAAAAA 3434	
Db	3999	AAACATGACTAA 4010	
RESULT 2			
AC019335/c			
LOCUS Homo sapiens chromosome 5 clone RP11-427K3, WORKING DRAFT SEQUENCE,			
DEFINITION 18 unordered pieces.			
AC019335			
AC019335.2 GI:7231064			
VERSION HTG; HTGS; PHASE1; HTGS_DRAFT.			
KEYWORDS Homo sapiens (human)			
SOURCE			
ORGANISM			
Homo sapiens			
REFERENCE			
1. (bases 1 to 182638)			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE			
1. The sequence of Homo sapiens clone			
Waterson, R.H.			
JOURNAL			
TITLE			
2. (bases 1 to 182638)			
REFERENCE			
Waterson, R.H.			
JOURNAL			
TITLE			
Direct Submission			
Submitted (01-JAN-2000) Genome Sequencing Center, Washington			
University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
MO 63109, USA			
On Mar 13, 2000 this sequence version replaced gi:6652510.			
COMMENT			
----- Genome Center -----			
Center: Washington University Genome Sequencing Center			
Center code: WUGSC			

```

Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0427K03
----- Summary Statistics -----
Sequencing vector: M13, 87%
Chemistry: Dye-terminator Big Dye, 13% of reads
Chemistry: Dye-terminator Big Dye, 13% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 174376 bases at least Q40
Consensus quality: 176799 bases at least Q30
Consensus quality: 178323 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 180938; sum-of-contigs
Quality coverage: 4.40 in Q20 bases; agarose-fp
Quality coverage: 4.46 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1193: contig of 1193 bp in length
* 1194 1293: gap of unknown length
* 1294 2928: contig of 1635 bp in length
* 2929 3028: gap of unknown length
* 3029 5126: contig of 2098 bp in length
* 5127 5226: gap of unknown length
* 5227 6713: contig of 1487 bp in length
* 6714 9443: gap of unknown length
* 9444 9543: contig of 2630 bp in length
* 9544 13186: gap of unknown length
* 13187 13287: contig of 3643 bp in length
* 13287 18101: gap of unknown length
* 18102 18201: contig of 4815 bp in length
* 18202 22856: gap of unknown length
* 22857 22957: gap of 4655 bp in length
* 22958 31608: contig of 8652 bp in length
* 31609 31708: gap of unknown length
* 31709 41973: contig of 10265 bp in length
* 41974 42074: gap of unknown length
* 42075 50868: contig of 8795 bp in length
* 50869 50969: gap of unknown length
* 50970 59200: contig of 8132 bp in length
* 59201 72609: gap of unknown length
* 72610 72709: contig of 13409 bp in length
* 72710 86964: gap of unknown length
* 86965 87064: contig of 14255 bp in length
* 87065 104522: gap of 17458 bp in length
* 104523 104623: gap of unknown length
* 104624 126460: contig of 21738 bp in length
* 126461 126461: gap of unknown length
* 126462 149770: contig of 23310 bp in length
* 149771 149870: gap of unknown length
* 149871 182638: contig of 32768 bp in length.
Location/Qualifiers
1. 182638
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-427K3"
1 1193
/misc_feature /note="assembly_name:Contig4"
1294 2928
/misc_feature /note="assembly_name:Contig5"
3029 5126
/misc_feature /note="assembly_name:Contig6"

```

```

misc_feature 5227..6713
/note="assembly_name:Contig7"
misc_feature 6814..9443
/note="assembly_name:Contig8"
misc_feature 9544..13186
/note="assembly_name:Contig9"
misc_feature 13287..18101
/note="assembly_name:Contig10"
misc_feature 18202..22856
/note="assembly_name:Contig11
vector_side:left"
misc_feature 22957..31608
/note="assembly_name:Contig12"
misc_feature 31709..41973
/note="assembly_name:Contig13"
misc_feature 42074..50868
/note="assembly_name:Contig14"
misc_feature 50969..59100
/note="assembly_name:Contig15"
misc_feature 59201..72609
/note="assembly_name:Contig16"
misc_feature 72710..86964
/note="assembly_name:Contig17
clone_end:right"
vector_side:right"
misc_feature 87065..104522
/note="assembly_name:Contig18"
misc_feature 104623..126360
/note="assembly_name:Contig19"
misc_feature 126461..149770
/note="assembly_name:Contig20"
misc_feature 149871..182638
/note="assembly_name:Contig21"
BASE COUNT 59178 a 34459 c 33797 g 53497 t 1707 others
ORIGIN
Query Match 59.0% Score 2045.4 DB 2; Length 182638;
Best Local Similarity 98.7% Pred. No. 0;
Matches 2072; Conservative 0; Mismatches 26; Indels 1; Gaps 1;
QY 1336 TGTCTTACCGTGGACCTTTCAGATGTCACCTTCACATGCGATGTCGCAAGAA 1395
Db 179790 TTTCTTTTCAGGAGCCTTCAGATGTCACCTTCACATGCGATGTCGCAAGAA 179731
QY 1396 TGGCAAGGTGACAAAGAGGAGAGCGGACATCTTTCGATGATCGCATCTAG 1455
Db 179730 TGGCAAGGTGACAAAGAGGAGAGCGGACATCTTTCGATGATCGCATCTAG 179671
QY 1456 CTGGACAGCAGCTGAGCGCAGCTCTCCCGAGAGAGCGTCACTGTGTCCCGTGC 1515
Db 179670 CTGGACAGCAGCTGAGCGCAGCTCTCCCGAGAGAGCGTCACTGTGTCCCGTGC 179611
QY 1516 ACGCAGACCAATCATCATCTGACCGTCCCATCAAGAGGCGCTGTCCAGCTGAT 1575
Db 179610 ACGCAGACCAATCATCATCTGACCGTCCCATCAAGAGGCGCTGTCCAGCTGAT 179551
QY 1576 CCAACAGTCAGAGTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1635
Db 179550 CCAACAGTCAGAGTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 179491
QY 1636 GGAATCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1695
Db 179490 GGAATCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 179431
QY 1696 GCTGAGAGAAATGATTCAGAGATTCCTTAAAGAGAGAGAGAGAGAGAGAGAGAG 1755
Db 179430 GCTGAGAGAAATGATTCAGAGATTCCTTAAAGAGAGAGAGAGAGAGAGAGAGAG 179371
QY 1756 AAAACATCTTGGCAATCTGAACTTTAAGAGAGATATCATCTTAAGAGAGAGAGAG 1815
Db 179370 AAAACATCTTGGCAATCTGAACTTTAAGAGAGATATCATCTTAAGAGAGAGAGAG 179311

```



QY 1816 CGGGGAAAAAAAAAGAGCTGATTTGAAATTAACCTCAATAAGAAATTCATATTT 1875  
 Db 179310 C-GGGGAAAAAAAAAGAGCTGATTTGAAATTAACCTCAATAAGAAATTCATATTT 179252  
 QY 1876 TAAAGAAAAAAATACACTAAATGATGACATTTCTTGAACAACAATATGATTCATATAT 1935  
 Db 179251 TAAAGAAAAAAATACACTAAATGATGACATTTCTTGAACAACAATATGATTCATATAT 179192  
 QY 1936 ACTACGCTCACTTAACTGATGACATTTCTTGAACAACAATATGATTCATATAT 1995  
 Db 179191 ACTACGCTCACTTAACTGATGACATTTCTTGAACAACAATATGATTCATATAT 179132  
 QY 1996 TAGATGTGAAGTACAAAGAAATCTTTTATGATTTATTTGTTTACTTGTCCATAT 2055  
 Db 179131 TAGATGTGAAGTACAAAGAAATCTTTTATGATTTATTTGTTTACTTGTCCATAT 179072  
 QY 2056 GCTAATATCTTAATATATGAGGACGATGATTAAGTATGATGATGATGATGATGAT 2115  
 Db 179071 GCTAATATCTTAATATATGAGGACGATGATTAAGTATGATGATGATGATGAT 179012  
 QY 2116 GCTAATATCTTAATATATGAGGACGATGATTAAGTATGATGATGATGATGAT 2175  
 Db 179011 GCTAATATCTTAATATATGAGGACGATGATTAAGTATGATGATGATGATGAT 178952  
 QY 2176 GTATGTAGAAAAAGATGCTGCGCAATGAGCAATGATGATGATGATGATGAT 2235  
 Db 178951 GTATGTAGAAAAAGATGCTGCGCAATGAGCAATGATGATGATGATGATGAT 178892  
 QY 2236 CTTTGTGTTATGAGGATGAGGAGGATGAGGATGATGATGATGATGATGATGAT 2295  
 Db 178891 CTTTGTGTTATGAGGATGAGGAGGATGAGGATGATGATGATGATGATGATGAT 178832  
 QY 2296 ATGTGCCAGTTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 2355  
 Db 178831 ATGTGCCAGTTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 178772  
 QY 2356 GTATGTAGAAAAAGATGCTGCGCAATGAGCAATGATGATGATGATGATGATGAT 2415  
 Db 178771 GTATGTAGAAAAAGATGCTGCGCAATGAGCAATGATGATGATGATGATGAT 178712  
 QY 2416 CTGCTCCAGACGCTCTCTTGAATATTTTCAATGATGATGATGATGATGAT 2475  
 Db 178711 CTGCTCCAGACGCTCTCTTGAATATTTTCAATGATGATGATGATGATGAT 178652  
 QY 2476 AAATTTCTTTTGGAGATTTCCAGGCTTTCTTCAACACATTTCTTCAACATGAT 2535  
 Db 178651 AAATTTCTTTTGGAGATTTCCAGGCTTTCTTCAACACATTTCTTCAACATGAT 178592  
 QY 2536 CTGATCATCAAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2595  
 Db 178591 CTGATCATCAAAATGATGATGATGATGATGATGATGATGATGATGATGAT 178532  
 QY 2596 TGCTTGGCTCTTTGCTGATTAAGAGAGATGAGAGAGCTGCTCAACATTTCTTGA 2655  
 Db 178531 TGCTTGGCTCTTTGCTGATTAAGAGAGATGAGAGAGCTGCTCAACATTTCTTGA 178472  
 QY 2656 TTAATAAAATCAAACTTCAATGATGATGATGATGATGATGATGATGATGAT 2715  
 Db 178471 TTAATAAAATCAAACTTCAATGATGATGATGATGATGATGATGATGATGAT 178412  
 QY 2716 TTCTCTACCTGAACTGAATATGATGATGATGATGATGATGATGATGATGAT 2775  
 Db 178411 TTCTCTACCTGAACTGAATATGATGATGATGATGATGATGATGATGATGAT 178352  
 QY 2776 ACACTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2835  
 Db 178351 ACACTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 178292  
 QY 2836 TGTGAACTTGAAGACACACAGATGATGATGATGATGATGATGATGATGATGAT 2895  
 Db 178291 TGTGAACTTGAAGACACACAGATGATGATGATGATGATGATGATGATGATGAT 178232  
 QY 2896 AGAAAGTTGATTCACACACAGATGATGATGATGATGATGATGATGATGATGAT 2955

Db 178231 AGAAAGTTGATTCACACACAGATGATGATGATGATGATGATGATGATGATGAT 178172  
 QY 2956 TACATTTTACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3015  
 Db 178171 TACATTTTACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 178112  
 QY 3016 CCCATGCTAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3075  
 Db 178111 CCCATGCTAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 178052  
 QY 3076 TGAAGCTCCAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3135  
 Db 178051 TGAAGCTCCAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 177992  
 QY 3136 TGAAGCTCCAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3195  
 Db 177991 TGAAGCTCCAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 177932  
 QY 3196 TATTTTACCTAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3255  
 Db 177931 TATTTTACCTAGATTTAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 177872  
 QY 3256 CAGTTTATTTTACGATTTCTCTAGATGATGATGATGATGATGATGATGATGATGAT 3315  
 Db 177871 CAGTTTATTTTACGATTTCTCTAGATGATGATGATGATGATGATGATGATGATGAT 177812  
 QY 3316 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3375  
 Db 177811 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 177752  
 QY 3376 GTGATTCACAGATTTAAAGCAATGATGATGATGATGATGATGATGATGATGAT 3434  
 Db 177751 GTGATTCACAGATTTAAAGCAATGATGATGATGATGATGATGATGATGATGAT 177693

RESULT 3  
 AC008716/c 184589 bp DNA linear PRI 17-OCT-2001  
 LOCUS Homo sapiens chromosome 5 clone C7B-85P21, complete sequence.  
 DEFINITION AC008716  
 ACCESSION AC008716.7 GI:16195190  
 KEYWORDS HTG.  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 REFERENCE  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 REFERENCE  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 COMMENT  
 On Oct 17, 2001 this sequence version replaced gi:8122137.  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing completed at Stanford Human Genome Center  
 www.hgsc.stanford.edu  
 Quality: Phrap Quality >=40 99.5% of Sequence;  
 Estimated Total Number of Errors is 0.9.



Db 177203 CAGTTTATTTTTCAGATTCCTCATGATTTTCAGTGTAAACCAAAAATATTTGTCAT 177144  
Qy 3316 TAAATGTTGGCCCAAGACCTCTAATTTGTTTATGCGGTGGTGCATGGTGT 3375  
Db 177143 TAAATGTTGGCCCAAGACCTCTAATTTGTTTATGCGGTGGTGCATGGTGT 177084  
Qy 3376 GTGTATCAGGTAAATTAAGGCAATGGATGATTTAAAAAATTTAAAAA 3434  
Db 177083 GTGTATCAGGTAAATTAAGGCAATGGATGATTTAAAAAATTTAAAAA 177025

RESULT 4  
AC008473/c 98360 bp DNA linear PRI 03-OCT-2001  
LOCUS Homo sapiens chromosome 5 clone CTC-375J15, complete sequence.  
DEFINITION AC008473  
ACCESSION AC008473  
VERSION AC008473.6 GI:15887240  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 98360)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
DIRECT SUBMISSION  
2 (bases 1 to 98360)  
DOE Joint Genome Institute.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
DIRECT SUBMISSION  
3 (bases 1 to 98360)  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
DIRECT SUBMISSION  
4 (bases 1 to 98360)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
DIRECT SUBMISSION  
Submitted (01-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 98360)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Oct 3, 2001 this sequence version replaced gi:3958005.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phred Quality >=40 99.1% of Sequence;  
Estimated Total Number of Errors is 0.8.  
STS Content:  
SHGC-103102 G57424.  
Location/Qualifiers  
1. 98360  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-375J15"  
BASE COUNT 31520 a 19442 c 18516 g 28882 t  
ORIGIN

Query Match 58.4% Score 2038.6; DB 9; Length 98360;  
Best Local Similarity 98.5%; Pred. No. 0;  
Matches 2068; Conservative 0; Mismatches 29; Indels 2; Gaps 1;

Qy 1336 TGTCTTCAACGTGAGCTTCAGATGTCACCTTCACATGGAGTTGCTGCAAGA 1395  
Db 6692 TTTCTTTTCAGGTGAGCTTCAGATGTCACCTTCACATGGAGTTGCTGCAAGA 6693  
Qy 1396 TGGCAAGGTGACAAAGAGGAGAGCGGACGCTTTGCAAGACCTTCACATCTAG 1455  
Db 6632 TGGCAAGGTGACAAAGAGGAGAGCGGACGCTTTGCAAGACCTTCACATCTAG 6573  
Qy 1456 CTGCGACAGCCAGTCTAGGCGCAGCTCTCCCGACGAGACGCTATCTGTGTCCTGAC 1515

Db 6572 CTGCGACAGCCAGTCTAGGCGCAGCTCTCCCGACGAGACGCTATCTGTGTCCTGAC 6513  
Qy 1516 AGCCGACCAACATCCAGACCTCTGAGACCGTCCATCAAGAGGCGCTGTCAGCTAT 1575  
Db 6512 AGCCGACCAACATCCAGACCTCTGAGACCGTCCATCAAGAGGCGCTGTCAGCTAT 6453  
Qy 1576 CCAAGAGTGAAGATGGGCGGAGAAAGGCACTTACCTCGAGTTCTGACTTCAGGCTCAG 1635  
Db 6452 CCAAGAGTGAAGATGGGCGGAGAAAGGCACTTACCTCGAGTTCTGACTTCAGGCTCAG 6393  
Qy 1636 GGAATGGAACATGAGCGACCAAAAAAAGCTGTCTAAAGAAAGCTCAATGAGAGGA 1695  
Db 6392 GGAATGGAACATGAGCGACCAAAAAAAGCTGTCTAAAGAAAGCTCAATGAGAGGA 6333  
Qy 1696 GCTGGAAGATGATCCAGATTTTCTTAAAAAATTTCCAGATCGGTTCTGAGAG 1755  
Db 6332 GCTGGAAGATGATCCAGATTTTCTTAAAAAATTTCCAGATCGGTTCTGAGAG 6273  
Qy 1756 AAAACATCTTGGCAATCTGAACTTTAAAGAAATGATCATATTAAGGAGGCTGGGG 1815  
Db 6272 AAAACATCTTGGCAATCTGAACTTTAAAGAAATGATCATATTAAGGAGGCTGGGG 6213  
Qy 1816 CCGGAAAAAATTTAAAAAGATCATTTGAAATTTAACTTAAAAAGATTCATATT 1875  
Db 6212 CCGG-AAAAAATTTAAAAAGATCATTTGAAATTTAACTTAAAAAGATTCATATT 6155  
Qy 1876 TAAAGAAAAAATTAACAATATGAGCAATTTCTTGAACAACAATGATCATATGAT 1935  
Db 6154 TAAAGAAAAAATTAACAATATGAGCAATTTCTTGAACAACAATGATCATATGAT 6095  
Qy 1936 ACTACGCTCTTAACTGATGATCACTTAACATGATTAATCAAGGAGTGTATTTCTTC 1995  
Db 6094 ACTACGCTCTTAACTGATGATCACTTAACATGATTAATCAAGGAGTGTATTTCTTC 6035  
Qy 1996 TAGATGGAAGTCAAGAAATCTTTTATGATTTGTTTCTTCTGCTCCATCT 2055  
Db 6034 TAGATGGAAGTCAAGAAATCTTTTATGATTTGTTTCTTCTGCTCCATCT 5975  
Qy 2056 GCTAATCTTATATATATATGAGAGGCACTAAGTAAAGTACGTAAGGCTTGGGA 2115  
Db 5974 GCTAATCTTATATATATATGAGAGGCACTAAGTAAAGTACGTAAGGCTTGGGA 5915  
Qy 2116 GTCAATTTATCCAACTGGGTTTTTCTCTCATCTTCTACCTCTCTTGAATGAG 2175  
Db 5914 GTCAATTTATCCAACTGGGTTTTTCTCTCATCTTCTACCTCTCTTGAATGAG 5855  
Qy 2176 GTATGGAAGAAAGATCTGGCCCAATGAGCATTAAGTTGAATTTTATTTGTTT 2235  
Db 5854 GTATGGAAGAAAGATCTGGCCCAATGAGCATTAAGTTGAATTTTATTTGTTT 5795  
Qy 2236 CTTTGTCTTATGAGGTTGGGGAAGGAGATTTATATGACTTTTCACTCAAACTAT 2295  
Db 5794 CTTTGTCTTATGAGGTTGGGGAAGGAGATTTATATGACTTTTCACTCAAACTAT 5735  
Qy 2296 ATGTGCAAGTTTATATGATCTCGTATGATGATTTTGTGCAACAAAGCAACTTA 2355  
Db 5734 ATGTGCAAGTTTATATGATCTCGTATGATGATTTTGTGCAACAAAGCAACTTA 5675  
Qy 2356 GTATGATATATACATGACGACATGATGAGGAGGCTTGAAGGCTGCTGCT 2415  
Db 5674 GTATGATATATACATGACGACATGATGAGGAGGCTTGAAGGCTGCTGCT 5615  
Qy 2416 CTGCTCCAGACCCCTCTTGAATATTTTCAAGATGATGAGCTTCTGACTTTCTTA 2475  
Db 5614 CTGCTCCAGACCCCTCTTGAATATTTTCAAGATGATGAGCTTCTGACTTTCTTA 5555  
Qy 2476 AAATCTTTTGGGAAATTTCCAGGCTTTTCAACAACAATTTTAAATCAATGACT 2535  
Db 5554 AAATCTTTTGGGAAATTTCCAGGCTTTTCAACAACAATTTTAAATCAATGACT 5495  
Qy 2536 CTGATCATCAAAATTTGATCTTATGAAATTAATTAACCTCAGGCTCATTTCAC 2595

Db	5494	TCATCATCAACAATGTATTCCTTAATGTGAATTAATACCTCGAGGTCATATTAC	5455
Qy	2596	TGCTTGCCCTTTGTCTGCTAGTAAGAGAGATAGAGAGAGTGCTCAACATTCCTGTG	2655
Db	5434	TGCTTGCTCTTTGTCTGCTATTAAGAGAGATAGAGAGAGTGCTCAACATTCCTGTG	5375
Qy	2656	TTAAAAAATTCAAACATTCATATCCAAATTTTCGTAAATGAATCCATCAGCTCAGC	2715
Db	5374	TTAAAAAATTCAAACATTCATATCCAAATTTTCGTAAATGAATCCATCAGCTCAGC	5315
Qy	2716	TTCTTAACCTGAAGCTGAATTAATACCTCTTTCTCAATGTTTCAAGGTTCTTACGCCC	2775
Db	5314	TTCTTAACCTGAAGCTGAATTAATACCTCTTTCTCAATGTTTCAAGGTTCTTACGCCC	5255
Qy	2776	ACAGTTTAATGTGTGTGCTTTCCACATTAATCCATTAAGTTCTGTGTCTGTGTGT	2835
Db	5254	ACAGTTTAATGTGTGTGCTTTCCACATTAATCCATTAAGTTCTGTGTGTGTGTGT	5195
Qy	2836	TGTGAAGCTAAGGACACACACAGTACTTGAATTAAGGTCGGGCTTTTGTTGTTAG	2895
Db	5194	TGTGAAGCTAAGGACACACACAGTACTTGAATTAAGGTCGGGCTTTTGTTGTTAG	5135
Qy	2896	AGAAAGTTGATTCACACACACACCTTAATTTCTTATAAAAATTTTAACACAAAGC	2955
Db	5134	AGAAAGTTGATTCACACACACACCTTAATTTCTTATAAAAATTTTAACACAAAGC	5075
Qy	2956	TACATTTTACTGTGTGTGAGCCGTTTTGTGCTTTGGGATTCGGGCTTTGGCTGTG	3015
Db	5074	TACATTTTACTGTGTGTGAGCCGTTTTGTGCTTTGGGATTCGGGCTTTGGCTGTG	5015
Qy	3016	CCCATGCTAGAGATTAGCTGTGTGATTTTAATGATCTGTACACACCAAGGTAAC	3075
Db	5014	CCCATGCTAGAGATTAGCTGTGTGATTTTAATGATCTGTACACACCAAGGTAAC	4955
Qy	3076	TGAAGCTCCAGATTAAGTTTCAGATTTCTAAATGAATCTTTTCAATTAATACC	3135
Db	4954	TGAAGCTCCAGATTAAGTTTCAGATTTCTAAATGAATCTTTTCAATTAATACC	4895
Qy	3136	TGACTTGTAAGACACAGCCAAAAGAACTGTTAATAGCATCCGTCTCATGTAATCTG	3195
Db	4894	TGACTTGTAAGACACAGCCAAAAGAACTGTTAATAGCATCCGTCTCATGTAATCTG	4835
Qy	3196	TATTTTACTAAGTACCAATAGCTCTTTTCATAGACTTGTGTCACAAAGGTTAAAAAGC	3255
Db	4834	TATTTTACTAAGTACCAATAGCTCTTTTCATAGACTTGTGTCACAAAGGTTAAAAAGC	4775
Qy	3256	CAGTTTAAATTTAGCATTCCTCATGATCTTCAAGTGTAAACCAAAATTAATGTCAAT	3315
Db	4774	CAGTTTAAATTTAGCATTCCTCATGATCTTCAAGTGTAAACCAAAATTAATGTCAAT	4715
Qy	3316	TAAATGTTGTGTGCCAAGACATCCTTAATTTGTTTAAATTCGTGTGTGATGTGTAT	3375
Db	4714	TAAATGTTGTGTGCCAAGACATCCTTAATTTGTTTAAATTCGTGTGTGATGTGTAT	4655
Qy	3376	GTGTATCAGCAGGTAAATTAAGGCAATGGAATGAATTAATTAATTAATTAATTAATTA	3434
Db	4654	GTGTATCAGCAGGTAAATTAAGGCAATGGAATGAATTAATTAATTAATTAATTAATTA	4596
RESULT 5			
LOCUS	AX405760	2412 bp	DNA
DEFINITION	Sequence 175 from Patent W00222660.		
ACCESSION	AX405760		
VERSION	AX405760.1		
KEYWORDS	GI:21438959		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1. Tang, Y. T., Izu, C., Zhou, P., Aundri, V., Zhang, J., Zhao, Q. A., Ren, F., Xue, A. J., Yang, Y., Weinman, T. and Dimauro, R. T.		

TITLE		Novel nucleic acids and polypeptides	
PATENT		Patent: WO 0222660-A,1175 21-MAR-2002;	
JOURNAL		HYSEO, INC. (US)	
FEATURES		Location/Qualifiers	
source		1..2412	
		/organism="Homo sapiens"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9606"	
		1092..2378	
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="CAD34867.1"	
		/db_xref="GI:21438960"	
		/translation="MALSGKCSRYPRDGSADVNSPEVELNNGGVYFTRHSLLT SIFSLHAKMSPKPDPTANDLADSKGSPRIIDRQSPFLIYLADLRDQVLPDHP KQSLKHEMLQVLDLYKLTLPDIKSPGSGSPEDSASQSDPIKPESSLLPAL KMGKFLITGKRSCTLDREQDAKAKRNPRLVLCGRISLAKVEPTEINSDPRDR BEFTYSRFLFKHLERAFMDSECEGFHVAACNSYRASPINQYMDPKTWSYEVV YRPSRSMHSDHCCCKNGKDGESGSCNDSJTSCHSOSEASPEYLCGPVLT RQNTIQDIPKIKGPVQLIQSEMRKRSQDLRLTLYTSKESMSKKAIVAKLSITE ELEKELIQDFLKIKIPRPFRRHNSDELKRYHL"	
BASE COUNT		638 a 585 c 551 g 638 t	
ORIGIN			
Query Match		47.3%; Score 1640.8; DB 6; Length 2412;	
Best Local Similarity		99.6%; Pred. No. 0;	
Matches 1655; Conservative		0; Mismatches 7; Indels 0; Gaps 0;	
Oy	183	ACCAATACGACATCTGATGTAAGTGGGGAATGGCCCTGCTGTGCATGTGAGCTTGATGGA	242
Db	760	ACTAAGACGGACATCTGATGTAAGTGGGGAATGGCCCTGCTGTGCATGTGAGCTTGATGGA	819
Oy	243	AGATTGGATATTAACAGAGTGTATATATTTTATGAAGTGGAGCGACTACATCCATCACA	302
Db	820	AGATTGGATATTAACAGAGTGTATATATTTTATGAAGTGGAGCGACTACATCCATCACA	879
Oy	303	TCCAGGGGTTAACTACTTTTACAGATCACTCACTGTGACCTCTTAATACATTTTGAT	362
Db	880	TCCAGGGGTTAACTACTTTTACAGATCACTCACTGTGACCTCTTAATACATTTTGAT	939
Oy	363	TTCCTGGGGGAAAAAATCTGGGATTAAGAGAGGTCAATTTTAAATAGATTAGATCCTT	422
Db	940	TTCCTGGGGGAAAAAATCTGGGATTAAGAGAGGTCAATTTTAAATAGATTAGATCCTT	999
Oy	423	TCCCTTCTTACAGTTGATCCAAAGATAGAGCTGTGACTCCATTTGATTCACCTTTA	482
Db	1000	TCCCTTCTTACAGTTGATCCAAAGATAGAGCTGTGACTCCATTTGATTCACCTTTA	1055
Oy	483	AATCAAAATAGCAGCAGCAGAAAGAAAGGACAATGCTCTGATGGAAAACTGATGCTGT	542
Db	1060	AATCAAAATAGCAGCAGCAGAAAGAAAGGACAATGCTCTGATGGAAAACTGATGCTGT	1111
Oy	543	ATTATCTCTCAGAAACAAGGGTCCGACGTTCCCAACTCTTCCCTGAGGTGGTAGACCTGA	602
Db	1120	ATTATCTCTCAGAAACAAGGGTCCGACGTTCCCAACTCTTCCCTGAGGTGGTAGACCTGA	1177
Oy	603	ATGTGGGGGGTCAAGTTTATTTTATCTGCCCATTCACAATTGATTAAGCATCCCTCATCC	662
Db	1180	ATGTGGGGGGTCAAGTTTATTTTATCTGCCCATTCACAATTGATTAAGCATCCCTCATCC	1233
Oy	663	TCCGTGGAAAAAGTTTTTCCCAAGAGAGACACGGCTATATATCTACCCAAAGACCTCCA	722
Db	1240	TCCGTGGAAAAAGTTTTTCCCAAGAGAGACACGGCTATATATCTACCCAAAGACCTCCA	1299
Oy	723	AGGAAAGTTTTTCAATGACAGATGATATCTTTGTCGGTATATATTCGTGACATATCTCA	782
Db	1300	AGGAAAGTTTTTCAATGACAGATGATATCTTTGTCGGTATATATTCGTGACATATCTCA	1355
Oy	783	GGGACAGGACAGGTGCTGCTGCTGATCACTTTCCAGAAAAAGGAAGCTGAAAAAGGAAAG	842
Db	1360	GGGACAGGACAGGTGCTGCTGCTGATCACTTTCCAGAAAAAGGAAGCTGAAAAAGGAAAG	1411
Oy	843	CTGAATTACTTCACAGCTCCACAGACTTGCTCAAACTCTGTACCCCCGATGAATTCAGCAAA	902

```

Db      1420  CTGATATCTTCCAGCTCCAGACTTGGTCAAACTCTGAGCCCCGAGTAATCAAGGAA 1479
Qy      903    GCCCAGATGATTTCTGCCCAAGTGACTTTGAAGATCCCTCCCAAGAAAGCAACAAGAA 962
Db      1480    GCCCAGATGATTTCTGCCCAAGTGACTTTGAAGATCCCTCCCAAGAAAGCAACAAGAA 1539
Qy      963    TCTGCCCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1022
Db      1540    TCTGCCCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1599
Qy      1023    ACAGAGATCTGCACTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1082
Db      1600    ACAGAGATCTGCACTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
Qy      1083    CCGGATTTTGGTTTGGTGGAGAGATTTCTTGGCAAAAAGATCTTTGGAGAACTTTGA 1142
Db      1660    CCGGATTTTGGTTTGGTGGAGAGATTTCTTGGCAAAAAGATCTTTGGAGAACTTTGA 1719
Qy      1143    ATGAAAGCAGAGACCTGATTCGAGCCCCAGAAAGATACCTCCAGATTTTATCTCAAT 1202
Db      1720    ATGAAAGCAGAGACCTGATTCGAGCCCCAGAAAGATACCTCCAGATTTTATCTCAAT 1779
Qy      1203    TCAAGCACTGGAAGGCTTTTGAATGTGTTCAGAGTGTGATTTCACTGAGGCTT 1262
Db      1780    TCAAGCACTGGAAGGCTTTTGAATGTGTTCAGAGTGTGATTTCACTGAGGCTT 1839
Qy      1263    GTAACTGATGATGACAGATCTTTCACTCAACCAATATACAGATGACAGATCTGATCA 1322
Db      1840    GTAACTGATGATGACAGATCTTTCACTCAACCAATATACAGATGACAGATCTGATCA 1899
Qy      1323    GCTACACTGATATGTCTTCTTACACGTGAGCTTTCAGATGTCTACCTCACTGAGAT 1382
Db      1900    GCTACACTGATATGTCTTCTTACACGTGAGCTTTCAGATGTCTACCTCACTGAGAT 1959
Qy      1383    GCTGCTGCAAGAGATGCAAGGTGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1442
Db      1960    GCTGCTGCAAGAGATGCAAGGTGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2019
Qy      1443    TCTCCACATCTAGCTGCGACAGCCAGCTGAGCGACGCTCTCCAGAGAGAGCTACCT 1502
Db      2020    TCTCCACATCTAGCTGCGACAGCCAGCTGAGCGACGCTCTCCAGAGAGAGCTACCT 2079
Qy      1503    GTGCTCCCTGAGACAGCCAGACCAACATCTCAACCTGAGACCTTCCCAAGAGAGGCT 1562
Db      2080    GTGCTCCCTGAGACAGCCAGACCAACATCTCAACCTGAGACCTTCCCAAGAGAGGCT 2139
Qy      1563    CTGTCACTGATTCCAACAGTCAAGAGATGCGGCGGAAAGCACTTACTCCGATTTCTGA 1622
Db      2140    CTGTCACTGATTCCAACAGTCAAGAGATGCGGCGGAAAGCACTTACTCCGATTTCTGA 2199
Qy      1623    CTTCAAGCTTCCAGAGATTCGAACATGAGACAGCAAAAAAAGCTTTTAAAGAAAGCTCT 1682
Db      2200    CTTCAAGCTTCCAGAGATTCGAACATGAGACAGCAAAAAAAGCTTTTAAAGAAAGCTCT 2259
Qy      1683    CAATTGAGAGAGAGCTGAGAAATGTATCCAGAGATTTCTTAAAAAAAATTCAGATC 1742
Db      2260    CAATTGAGAGAGAGCTGAGAAATGTATCCAGGATTTCTTAAAAAAATTCAGATC 2319
Qy      1743    GGTTCCTGAGAGAAACATCCTGGGCACTGAACTTTTAAAGAAAGATCATTAAG 1802
Db      2320    GGTTCCTGAGAGAAACATCCTGGGCACTGAACTTTTAAAGAAAGATCATTAAG 2379
Qy      1803    GGAAGGCTGGGGCGGCGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1834
Db      2380    GGAAGGCTGGGGCGGCGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2411

```

RESULT 6  
 AC008652/c AC008652 171949 bp DNA 11near PRI 31-JUL-2001  
 LOCUS Homo sapiens chromosome 5 clone CTB-18F1, complete sequence.  
 DEFINITION  
 AC008652  
 ACCESSION

```

VERSION AC008652.6 GI:15042788
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE (bases 1 to 171949)
JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE Direct Submission
AUTHORS Unpublished
TITLE 2 (bases 1 to 171949)
JOURNAL DOE Joint Genome Institute.
REFERENCE Direct Submission
AUTHORS Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 171949)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE Submitted (31-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 31, 2001 this sequence version replaced gi:9256021.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
FEATURES
source
1..171949
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-18F1"
BASE COUNT 54666 a 33484 c 31658 g 52141 t
ORIGIN
Query Match 33.5%; Score 1161.8; DB 9; Length 171949;
Best Local Similarity 99.8%; Pred. No. 1.4e-240;
Matches 1163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 187 ATACGACATCTGATGAACTGGGAAATTTGGCTGCTGATGAGCTGATGAGAAAT 246
Db 171266 AGACGACATCTGATGAACTGGGAAATTTGGCTGCTGATGAGCTGATGAGAAAT 171207
Qy 247 TGAATATGACAGATGATATATATATATATATATATATATATATATATATATATAT 306
Db 171206 TGAATATGACAGATGATATATATATATATATATATATATATATATATATATAT 171147
Qy 307 GGGTTTAACTACTTTTCAAGCATCACTTCACTGTGACCTTATACATTTTGAATTTCT 366
Db 171146 GGGTTTAACTACTTTTCAAGCATCACTTCACTGTGACCTTATACATTTTGAATTTCT 171087
Qy 367 TGGGGGAAAAATATAGGATATAGAGAGGTCATTTTATATAGTTAGCAATCTTTTCC 426
Db 171086 TGGGGGAAAAATATAGGATATAGAGAGGTCATTTTATATAGTTAGCAATCTTTTCC 171027
Qy 427 TTCTTACAGATGATTCGAAGATATAGGCTGATGATGATGATGATGATGATGATGATGAT 486
Db 171026 TTCTTACAGATGATTCGAAGATATAGGCTGATGATGATGATGATGATGATGATGATGAT 170967
Qy 487 AAATATGACGACGAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAG 546
Db 170966 AAATATGACGACGAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAG 170907
Qy 547 TCCTGAGAAACAAGGTCGAGATTCGAACTCTTCTCTGAGTGTGAGAGCTGAATGT 606
Db 170906 TCCTGAGAAACAAGGTCGAGATTCGAACTCTTCTCTGAGTGTGAGAGCTGAATGT 170847
Qy 607 CGGGGCTCAAGTTTATTTTACTCGCCATTCGACATTTGATTAAGCATTCCTCATTCCT 666
Db 170846 CGGGGCTCAAGTTTATTTTACTCGCCATTCGACATTTGATTAAGCATTCCTCATTCCT 170787

```

```

QY 667 GTGGAATATGTTTCCCAAGAGACAGCGTAATGATCTAGCCAGAGCTCCAGG 726
DB 170786 GTGGAATATGTTTCCCAAGAGACAGCGTAATGATCTAGCCAGAGCTCCAGG 170727
QY 727 AAGTTTTCATGACAGAGATGATCTGTTCCGTTATATTCCTGAGCTATCCAGG 786
DB 170726 AAGTTTTCATGACAGAGATGATCTGTTCCGTTATATTCCTGAGCTATCCAGG 170667
QY 787 CAGGAGAGTGTCTGCTGCTGATCACTTCCAGAAAAGAAACAGAAAAGAGAGCTGA 846
DB 170666 CAGGAGAGTGTCTGCTGCTGATCACTTCCAGAAAAGAAACAGAAAAGAGAGCTGA 170667
QY 847 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 906
DB 170606 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 170547
QY 907 AGATGAATTCGACAGAGTGAATGAGTCTCCAGAGAAAGAGAGAGAGAGATCTG 966
DB 170546 AGATGAATTCGACAGAGTGAATGAGTCTCCAGAGAAAGAGAGAGAGAGATCTG 170487
QY 967 CCCCCCTTCTCCCTGCTCTCCGACCGAGAGTGGGTTTCACTTACTGTGAGTTA 1026
DB 170486 CCCCCCTTCTCCCTGCTCTCCGACCGAGAGTGGGTTTCACTTACTGTGAGTTA 170427
QY 1027 AGATTCCTGACCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
DB 170426 AGATTCCTGACCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170387
QY 1087 GATTTCGTTTGTGAGAGATTCCTTGGCAAAAGAGCTTGGAGAACTTGAATGA 1146
DB 170386 GATTTCGTTTGTGAGAGATTCCTTGGCAAAAGAGCTTGGAGAACTTGAATGA 170387
QY 1147 AAGGAGAGACCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
DB 170306 AAGGAGAGACCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170247
QY 1207 GCACCTGAAAAGGCTTTGATATGTTTCAAGAGTGTGATTCACATGTGCTGTA 1266
DB 170246 GCACCTGAAAAGGCTTTGATATGTTTCAAGAGTGTGATTCACATGTGCTGTA 170187
QY 1267 CTCAATGAGTGAAGATCTTTCATCAACCAATATACAGATCAAGATCTGTAAGTA 1326
DB 170186 CTCAATGAGTGAAGATCTTTCATCAACCAATATACAGATCAAGATCTGTAAGTA 170127
QY 1327 CACTGAATATGCTCTTACCGTAG 1351
DB 170126 CACTGAATATGCTCTTACCGTAG 170102

```

```

COMMENT
Drive, Walnut Creek, CA 94598, USA
On May 1, 2001 this sequence version replaced gi:13699337.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.8.
FEATURES
source
1..209114
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-222022"
BASE COUNT 65378 a 39329 c 40172 g 64235 t
ORIGIN
Query Match 33.5%; Score 1161.8; DB 9; Length 209114;
Best Local Similarity 99.8%; Pred. No. 1.4e-240;
Matches 1163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
187 ATACGACATCTAGTACTGGGAAATGGCTGCTGCTCATGTGAGCTTGATGAAGAT 246
168231 AGACGACATCTAGTACTGGGAAATGGCTGCTGCTCATGTGAGCTTGATGAAGAT 168290
QY 247 TGAATATGACAGATGATTAATTAATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 306
DB 168291 TGAATATGACAGATGATTAATTAATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 168350
QY 307 GGGTTTAATCACTTTTTCAGATCACTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 366
DB 168351 GGGTTTAATCACTTTTTCAGATCACTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 168410
QY 367 TGGGGGAAAAATCTGGATTAAGAGAGAGATTTTAAATAGTACATCTTTTCC 426
DB 168411 TGGGGGAAAAATCTGGATTAAGAGAGAGATTTTAAATAGTACATCTTTTCC 168470
QY 427 TTTCTTCAAGATTGACCAAGAGATGAGCTGAGACTCAATGATGACCTTTAAATC 486
DB 168471 TTTCTTCAAGATTGACCAAGAGATGAGCTGAGACTCAATGATGACCTTTAAATC 168530
QY 487 AAAATAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546
DB 168531 AAAATAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 168590
QY 547 TCCTCGAGAAACAAGAGTCCGAGTCCCAATCTCTTCCCTGAGTGTGAGCTGAATGT 606
DB 168591 TCCTCGAGAAACAAGAGTCCGAGTCCCAATCTCTTCCCTGAGTGTGAGCTGAATGT 168650
QY 607 CGGGGGTCAAGTTTATTTTACTCGCATTCACATGATAGATCCCTCATTCCTCT 666
DB 168651 CGGGGGTCAAGTTTATTTTACTCGCATTCACATGATAGATCCCTCATTCCTCT 168710
QY 667 GTGGAATATGTTTCCCAAGAGACAGCGTAATGATCTAGCCAGAGCTCCAGG 726
DB 168711 GTGGAATATGTTTCCCAAGAGACAGCGTAATGATCTAGCCAGAGCTCCAGG 168770
QY 727 AAGTTTTCATGACAGAGATGATCTGTTCCGTTATATTCCTGAGCTATCCAGG 786
DB 168771 AAGTTTTCATGACAGAGATGATCTGTTCCGTTATATTCCTGAGCTATCCAGG 168830
QY 787 CAGGAGAGTGTCTGCTGCTGATCACTTCCAGAAAAGAAAGCTGAAGAGAGCTGA 846
DB 168831 CAGGAGAGTGTCTGCTGCTGATCACTTCCAGAAAAGAAAGCTGAAGAGAGCTGA 168890
QY 847 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 906
DB 168891 ATACTTCAGCTCCAGACTGTGTCAAACTCTGACCCCGATGAATCAAGCAAGCCC 168950
QY 907 AGATGAATTCGACAGAGTGAATGAGTCTCCAGAGAAAGAGAGAGAGAGATCTG 966
DB 168951 AGATGAATTCGACAGAGTGAATGAGTCTCCAGAGAAAGAGAGAGAGAGATCTG 169010

```

```

RESULT 7
AC008383 209114 bp DNA linear PRI 01-MAY-2001
LOCUS Homo sapiens chromosome 5 clone CTC-222022, complete sequence.
DEFINITION AC008383
AC008383
AC008383.8 GI:13899395
VERSION
KEYWORDS
HTG
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 209114)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 209114)
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 209114)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell

```



Qy	967	CCCCCTTCCTCCCTGCTCCCTCCGACCGCAAGGGGGGTTTCATTACCTGTCGGTTACG	1026
Db	169011	CCCCCTTCCTCCCTGCTCCCTCCGACCGCAAGGGGGGTTTCATTACCTGTCGGTTACG	169070
Qy	1027	AGGATCCTGCACCTTGGGCGAGAGGGAACGGCAGATGCCAAGTTTCGAGAGTTCCCG	1086
Db	169071	AGGATCCTGCACCTTGGGCGAGAGGGAACGGCAGATGCCAAGTTTCGAGAGTTCCCG	169130
Qy	1087	GATTTTGGTTTGGAGAGATTCCTTGCGCAAAAAGATCTTTGGAGAACTTTGAAATGA	1146
Db	169131	GATTTTGGTTTGGAGAGATTCCTTGCGCAAAAAGATCTTTGGAGAACTTTGAAATGA	169190
Qy	1147	AAGCAGAACCCCTGATGCGAACCCCGCAAGATACACCTCCGATTTTATCTCAATTCGA	1206
Db	169191	AAGCAGAACCCCTGATGCGAACCCCGCAAGATACACCTCCGATTTTATCTCAATTCGA	169250
Qy	1207	GCACCTGGAAAAGGCTTTGATATGTTGTGCAGAGTGGATTCGACATGGTCCTGTAA	1266
Db	169251	GCACCTGGAAAAGGCTTTGATATGTTGTGCAGAGTGGATTCGACATGGTCCTGTAA	169310
Qy	1267	CTCATCGGTGACAGCACTTTTATCAACCATATATACAGATGACAGATCTGGTCAAGCTA	1326
Db	169311	CTCATCGGTGACAGCACTTTTATCAACCATATATACAGATGACAGATCTGGTCAAGCTA	169370
Qy	1327	CAGTGATATATCTTTTACCGGTGAG	1351
Db	169371	CAGTGATATATCTTTTACCGGTGAG	169395

RESULT 8			
AC127249/c			
LOCUS	AC127249	135132 bp	DNA linear HTG-12-JUN-2003
DEFINITION	Mus musculus chromosome UNK clone RP24-475B8, WORKING DRAFT SEQUENCE, 4 unordered pieces.		
ACCESSION	AC127249		
VERSION	AC127249.3	GI:31621481	
KEYWORDS	HTGS PHRASE1; HTGS DRAFT; HTGS_ACTIVEFIN.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Wilson, R.K		
TITLE	1 (bases 1 to 135132)		
JOURNAL	The sequence of Mus musculus clone		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 135132)		
TITLE	McPherson, J.D. and Hesterston, R.H.		
JOURNAL	Direct Submission		
TITLE	Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
JOURNAL	3 (bases 1 to 135132)		
TITLE	Wilson, R.K.		
AUTHORS	Direct Submission		
TITLE	Submitted (12-JUN-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
JOURNAL	On Jun 12, 2003 this sequence version replaced GI:21886968.		
COMMENT			

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: M05SC
Web site: http://genome.wustl.edu
Contact: submissions@wuston.wustl.edu
Project Information
Center project name: M.BB0475808
-----
----- Summary Statistics -----
Sequencing vector: MJ3: 0%
Sequencing reagent: plasmid: 100%
Chemistry: Dye primer ET: 0% of reads
Chemistry: Dye-terminator: Big Dye: 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 133026 bases at least Q40
-----

```

```

Consensus quality: 133293 bases at least Q30
Consensus quality: 133464 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 8786: contig of 8786 bp in length
*
* 8687 8686: gap of unknown length
*
* 8687 8687: gap of unknown length
*
* 23581 23581: contig of 14659 bp in length
*
* 23582 23581: gap of unknown length
*
* 23582 23586: contig of 19933 bp in length
*
* 43516 43516: gap of unknown length
*
* 43517 133152: contig of 91416 bp in length.
*
* 43517 133152: contig of 91416 bp in length.
*
* LocalQualifiers
* 1 8786:132 133464:132
*
FEATURES
source

```

	misc_feature	1. .8786	/note="assembly_name:Contig21"	
	misc_feature	8887. .23581	/note="assembly_name:Contig22"	
	misc_feature	23682. .43616	/note="assembly_name:Contig23"	
	misc_feature	43717. .135132	/note="assembly_name:Contig24"	
BASE COUNT	44115 a	26731 c	25600 g	38364 t
ORIGIN			322 others	
Query Match	24.6%	Score 853.2;	DB: 2;	Length 135132;
Best Local Similarity	70.7%	Pred. No. 7.5e-174;		
Matches 14/3;	Conservative	0;	Mismatches 523;	Indels 86;
			Gaps	22;
Dy	1347	GTGAGCTTCCAGATGCTCAACCTTCACACCTGCGATTCTGCTGCAGAAATGGCAAAAGTG	1406	
Dy	124743	GTGAGCTTCCCGGTGCTCTCTCTCACTGTGACTGTGTCGAAAGATGGCAAGGAG	124684	
Oy	1407	ACAAAGAGGGAGAGGGACAGTCTTGSCAATGACCTCTTCACATCTAGCTGCAGACCC	1466	
Dy	124683	ACA--AAGAGAGAGCGGCACCTCCTCGAATGACCTGTCACTTCACCTGTCACACC	124622	
Oy	1467	AGTCGAGCGCAGCTCTCCACAGAGAGGTCATCTGTGTTCCCGTACACGCCAGACCA	1526	
Dy	124626	AGTCAGAGGCGCAGCTCTCCGACAGAGAGCGGTGATCTGTGGGCTGTGAAGGCCCCAGACCA	124567	
Oy	1527	ACATCCAGACTGTGACCGTCCATCAACAGAGGCGCTGTCCACGTGATCCAAACGTGAG	1586	
Dy	124566	ACATCCAGACTGTGATCCGCGCCATCAAGAAAGGTCCGATGACGTGATCCAAACGTGAG	124507	
Oy	1587	AGATCGCGCGGAAAAAGGACCTTACTCCGGAATTCGATTCAGGCTCCAGGAAATGAAACA	1646	
Dy	124506	AGATGAGCGCGGAAAAAGTGAACCTGCTCCGGAATTCGACGTGAGCTCCAGGAGTGAACA	124447	
Oy	1647	TGAGAGCGAAAAAAGAGCTGTTAAGAAAAAGCTCTCAATTGAGAGAGCTGAGAAAT	1706	
Dy	124446	TGAGAGCGAAAAAAGAAAGCTGCGAAGAAAAAGCTCTCAATCGAAGAAAGCTGAGAAAT	124387	
Oy	1707	GTATCCAGAGATTCTTAATAAAAAAATTCAGATCGGTTCTGTGAGAAAAATCTCTT	1766	
Dy	124386	GTATCCAGAGATTCTTGAAGATTAATAAATTCAGATCGGTTCTGTGAGAAAAATCTCTT	124327	
Oy	1767	GGCATCTGAACCTTTTAAGAAAGTATCATATTAAGGAGGCGTGGGGCGGGAAAAAA	1826	
Dy	124326	GCGAGCTGAACCTTTTAAGAAAGTATCATATTAAGGAGGCGGCTGTG-----	124278	
Oy	1827	AAAAAAAAGTCAATTTGAATTAACCTCAATAAAGAAATTCATATTTTAAGAAAAA	1886	





TITLE  
JOURNAL  
REFERENCES  
AUTHORS

Zainoun, J., Zemsek, L., Zimmer, A., and Zody, M.  
Direct Submission  
Submitted (11-MAR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 186417) [zainoun@mit.edu](mailto:zainoun@mit.edu) [zody@mit.edu](mailto:zody@mit.edu)

JILS  
JOURNAL  
COMMENT

Submitted 05-JUN-2003 Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 22, 2003 this sequence version replaced gi:30023906.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Whitehead Institute/ MIT Center for Genome Research  
Center code: WtBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: L19035  
Center clone name: 248\_F9

NOTE: this is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	32842: contig of 32842 bp in length
32843	32842: gap of 100 bp
32844	61050: contig of 29008 bp in length
61051	61050: gap of 100 bp
61951	70963: contig of 8913 bp in length
60561	71063: gap of 100 bp
70964	71063: contig of 2967 bp in length
71064	71430: gap of 100 bp
74031	74130: gap of 100 bp
74131	162009: contig of 88679 bp in length
16210	162009: gap of 100 bp
16211	186417: contig of 23508 bp in length

**FEATURES**  
**SOURCE**

```
/organism="Mus musculus"  
/mol_type="genomic DNA"  
/db_xref="taxon:10090"
```

BASE C  
ORIGIN

```

/clone_11b="RPCI-23 Female Mouse BAC"
54373 a 36475 c 36595 g 58359 t 615 others

```

Query Match	24.6%	Score 853.2;	DB 2;	Length 186417;
Best Local Similarity	70.7%;	Pred. No. 7.5e-174;		
Matches 1473; Conservative	0;	Mismatches 523;	Indels 86;	Gaps 22

Oy	1344	GTGACCTTCCAGATGGTCAACCTCACACTGCAATTCSTCTGCAAGAAATGGCAAAAGTG	1406
Db	172660	GTGACCTTCCCGGTGGTCTCTCTCATTTGATCTGCTCTGCAAGAAATGGCAAGGAG	172719
Oy	1407	ACAAAAGAGGGAGAGCGGCACGTCTTGSCAAATGACCTCTGCACATCTACTGCGACAGCC	1466
Db	172720	ACA---AAGAGAGAGGGAGACCTCTGGAATAGACCTGTCCACTTCACCTTGTGACAGCC	172776
Oy	1467	AGTCTGAGCGCAGCTCTCCCGAGAGACGTCATCTGTGTCCTCGTACACGCCAGACCA	1526
Db	172777	AGTCAGAGGCGACGCTCTCCGACAGAGAGCGGATCTGTGGGCTCTGAAAGCGCCAGACA	172838
Oy	1527	ACATCAGACTCTGSCACCGTCCCATCAAGAAAGGCGCTGTCCAGCTGATCCAAAGTCCAG	1586
Db	172837	ACATCAGACTCTGAGATCCGGCCCATCAAGAAAGGTCGGGTGACAGCTGATCCAAAGTCCAG	172896
Oy	1587	AGATCGCGCGGAAAAAGCATTAATCCGAGATTCTGACTTCAAGGCTCAGGGAAATCGAAACA	1646
Db	172897	AGATGAGCGGAAAAAGTACGTCTCCGACTCTGACGTGAGCTTCAGGGATCGAGAAACA	172956
Oy	1647	TGAGAGCAAAAAAAGAGCTGTAAAGAAAGCTCTCAATGAGAGAGAGCTGAGAAAT	1706
Db	172957	TAGAGCAAAAAAAGAAAGCTCGAAGAGAAAGCTCTCATCGAGAGAAAGCTGAGAAAT	173016
Oy	1707	GTATCAGAGATTCTTAAAAAATAATTCAGATCGGTTTCTGAGAAACATCTTT	1766
Db	173017	GTATCAGAGATTCTTGGAAAGATAAAAATTCAGATCGCTTCTCTGAGCAAAACATCTTT	173076
Oy	1767	GGCAATCTGAACTTTTAAAGAAATGATCATCTATTAAGGAGGCGTGGGCGGGAAAAA	1826
Db	173077	GGCAATCTGAACTTTTACGAAAGTATCATATAGGGGAGGGCTGTGG-----	173125
Oy	1827	AAAAAAGAGTCAATTTTGAATTTAACTCTCAATAAAGAAATTCATATTTTAAAGAAAA	1886
Db	173126	----GTATCGCCACTTTGAATTAATACCTCCCAAGAGAAATATGTTAAAGSAAAA	173181
Oy	1887	AATACACTAATGATGCAKATTTCTTGAACAACAATGTCCATGTATATACATCGCTCA	1946
Db	173182	TA-ACAACTAAACGGTCCACATTTGTTAATGATACAAT-GTCCATTGATGATCACTGCTCA	173239
Oy	1947	CTTTCACCTAGTTCACCTTAACTGTAAATCCACAGGGTAGATTTCTTTAGATGTGAA	2006
Db	173240	CTTTCACCTAGTTCACCTTAACTGTGTAAATCCACAGGGTAGATTTCTTTAGATGTGAA	173299
Oy	2007	GTACAGAGAAATCTTTTATGATATTGG---TTTGTTTACTTGTCCTCAGTGTCAACTA	2063
Db	173300	CGAAGAAAGAGCTCTTATGTTGCTCTTGTATTTACTTGTCCTCAGTGTGTGAATA	173358
Oy	2064	TCTT-ATTAATAAGAGAGCGCATGACATAAAGTACGTGAGAGGCTCTGGAGATCATTT	2122
Db	173360	TCTTAAAGTACACAGAGACAGCTGACGTGAGATGACTCAGAGCTTTGGGAAATCATTTG	173419
Oy	2123	ATCCCAACGTGGTTTTT-----CTCTCATCCTTCTAC	2156
Db	173420	ATCCCAACCGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTCTCATTTTTTCTGAC	173479
Oy	2157	CTCCCTCTTTTGA---ATGAGGATAGTAGAAAAAATCTGGCCCAATGGCATAGTTTG	2214
Db	173480	CTCCCTCCCTTGAACCAAGATGAGCAGTTGAAAGAGATATAACCCGCTGCTTAGTTTAA	173539
Oy	2215	GAATTTTAAATTTTGGTTTTTCT---TTTGTTATAGGGTT---GGGCGGAATGCGACAT	2269
Db	173540	GAATTTATCTTTTCCCTTCTCTTTGTTTGTATATGGGGTGTAGGGGAGAAATGCGCAAT	173599
Oy	2270	TATATAGCTTTTACGCAAACTATAATAGTGCAGTTTATATTTAGATCTCGATATGATAG	2329
Db	173600	TGTATGATTTTTCATTAATAATCTCATATGTCCAGGTTCTATTTGACTTTTGTATGTGATAG	173659
Oy	2330	TATTTTGCAACACAGCA-CAACTAGATGATATATACATGAGCGCACAGATGCCAG	2388
Db	173660	CGTTTTCACACAGCAAGTATATGTCTGATATATGCAACAAAGATTCACAGACTTA	173719

```

Qy 2389 GGCCTAGACCTCCCAAGGGCTGTGCTCCTGCTCCCAAGACCCCTCTTGAATATTTCA 2448
    |||
Db 173720 GGGCCTGACAGAGAGGGCTTAATCTTACTATACAGTCCGCTTCAAGAGACCTTCA 173778
Qy 2449 GATGATGAGCTCTTACACTCTTCTTAAATCTTCTTGGGAAGATTTCCAGCCCTTCTT 2508
    |||
Db 173779 GACAAACAGAGCTCTGCTCTTATTTATTTAAACCTTCTGGGAGATTTCCAGCCCTTCTT 173838
Qy 2509 CACAACACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2563
    |||
Db 173839 GGGCAGGACAGCACTTCTTAAGCTGAATAGGCCCATCAATCAAAATTTCTTCTTAT 173898
Qy 2564 TGTGAATTAATTAACCTCAGGCTCAATTTACTGCTTCTGCTCTGATTAAGA 2623
    |||
Db 173899 TTTGAATTAATTAACCAAGCTCTCTTGAATTTATGTTCTTCTCTCAATTAAGAGC 173958
Qy 2624 GATGAGAGAGCTGCTCAAACTCTGCTGTTAA--AAATCAAACTCAATTC 2680
    |||
Db 173959 TGTGAATTAAGCTGCTCAAACTCTGCTGTTAA--AAATCAAACTCAATTC 174018
Qy 2681 ACMAAATTTCTCTTAATGACTCAACACTGCTTCTCTCACTGAACTGAATTTCA 2740
    |||
Db 174019 ACMAAAGAG--TGTGAAGAGTCTGACCTGCTCTCTCTCTCTCTCTCTCTCTCT 174077
Qy 2741 CCGTTTCTCTCAATGTTTCAAGATTTCTTACTGCTTCTCTCACTGAACTGAATTTCA 2800
    |||
Db 174078 CCGTTTCTCTCAATGTTTCAAGATTTCTTACTGCTTCTCTCACTGAACTGAATTTCA 174137
Qy 2801 CATATCCACATTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2860
    |||
Db 174138 TATATCCAGGCAAAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174197
Qy 2861 ACTGAATTAAGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2919
    |||
Db 174198 GATGAAATGCAATCCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174257
Qy 2920 CTAATTAATTTCTTAATAAATTTAACTTAACTAAGCACTAATTTTCTGCTGCTGCTGCT 2979
    |||
Db 174258 GCTTAATTTCTTAATAAATTTAACTTAACTAAGCACTAATTTTCTGCTGCTGCTGCT 174317
Qy 2980 TTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3038
    |||
Db 174318 GGTGATGATACCTTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174377
Qy 3039 CATTTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3098
    |||
Db 174378 AATTTCTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174437
Qy 3099 AGATTTCTAAATGAACTATCTTTTCAATTAATCACTGCTGCTGCTGCTGCTGCTGCT 3158
    |||
Db 174438 GATTTTCTAA--CTTTTAA--CTTTTAA--CTTTTAA--CTTTTAA--CTTTTAA-- 174488
Qy 3159 AAGAACTGTTAATAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3218
    |||
Db 174489 AAGAACTGTTAACA--CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174547
Qy 3219 TCTTTCATAGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3277
    |||
Db 174548 TCTTTCATAGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174606
Qy 3278 CATGATTTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3337
    |||
Db 174607 CATGATTTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174662
Qy 3338 CCTAATTTGTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3379
    |||
Db 174663 CCTAATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174704

```

RESULT 10  
 AC117867/c AC117867 242679 bp DNA linear HTG 11-OCT-2002  
 DEFINITION Rattus norvegicus clone CH230-37619, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 11 unordered pieces.  
 AC117867  
 HTG, HTGS PHASE1, HTGS DRAFT, HTGS ENRICHED.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 242679)  
 Muny, D., Marie, Metzger, M., Lee, Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D.,  
 Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Bialato, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Burch, P., Burrell, K., Caldwell, R.,  
 Cadenas, V., Carter, K., Cavazo, I., Casar, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, H., Davis, C., Day, Carroll, L., De Ande, C., Dedrich, D.,  
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, R., Divya, K.,  
 Dwyer, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Davies, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falis, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, J.,  
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,  
 Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,  
 Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kous, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lounsbury, L., Louie, H., Lozano, R., Lu, X., Ma, J.,  
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
 Mangum, B., Mapa, P., McNeill, T.Z., Meenen, E.,  
 Manthey, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
 Morgan, M., Morris, K., Morris, S., Muniasoon, A., Murphy, M., Neitz, L.,  
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
 Nwokwelen, O., Okunom, G., Olanpinu, A., Pal, S., Parks, K.,  
 Patel, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
 Pappas, M., Pappas, A., Popovic, D., Pritts, E., Pu, L.,  
 Puzos, F., Puzos, F., Puzos, F., Puzos, F., Puzos, F., Puzos, F.,  
 Rivas, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S.,  
 Sanders, W., Savary, G., Scheer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smale, D.,  
 Snel, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
 Steidle, M., Strong, R., Sutton, A., Swatek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F.,  
 Williams, G., Willison, R., Wiczek, R., Woodson, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, S., Dunn, D., von  
 Weinstock, G., and Gibbs, R.A.  
 Direct Submission  
 2 (bases 1 to 242679)  
 Morley, K. C.  
 Submitted (11-APR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 242679)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Oct 9, 2002 this sequence version replaced gi:21746224.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

## ----- Genome Center

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GT2A  
 Center clone name: CH230-37619

## ----- Summary Statistics

Assembly program: Phrap; version 0.990129  
 Consensus quality: 188097 bases at least Q40  
 Consensus quality: 190770 bases at least Q30  
 Consensus quality: 192614 bases at least Q20  
 Estimated insert size: 191086; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 11 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 185634: contig of 185634 bp in length  
 185635 185734: gap of unknown length  
 185735 189337: contig of 3603 bp in length  
 189338 189438: gap of unknown length  
 189439 198672: contig of 9235 bp in length  
 198673 198773: gap of unknown length  
 198774 202047: contig of 3275 bp in length  
 202048 202148: gap of unknown length  
 202149 203477: contig of 1330 bp in length  
 203478 203578: gap of unknown length  
 203579 204617: contig of 1040 bp in length  
 204618 204718: gap of unknown length  
 204719 205948: contig of 1131 bp in length  
 205949 207200: gap of unknown length  
 207201 207299: contig of 1251 bp in length  
 207300 208535: gap of unknown length  
 208536 211249: gap of unknown length  
 211250 211350: contig of 2614 bp in length  
 211351 242679: gap of unknown length  
 242680 242679: contig of 31330 bp in length.

## FEATURES

## SOURCE

1. 242679  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-37619"  
 misc\_feature  
 1 1668  
 /note="wgs end extension  
 clone\_end:596"  
 1914..2738  
 /note="clone boundary  
 clone\_end:596  
 site:Mbol

misc\_feature end\_sequence:RXAMW537V  
 8953..9899  
 /note="clone boundary  
 clone\_end:17  
 site:Mbol  
 misc\_feature end\_sequence:RXAMW537V  
 185735..187055  
 /note="wgs end extension  
 clone\_end:17"  
 misc\_feature 188311..189337  
 /note="wgs end extension  
 clone\_end:17"  
 misc\_feature 198773..199845  
 /note="wgs end extension  
 clone\_end:17"  
 BASE COUNT 59286 a 39026 c 39489 g 55897 t 48981 others  
 ORIGIN  
 Query Match 23.6%; Score 819.4; DB 2; Length 242679,  
 Best Local Similarity 83.8%; Pred. No. 1,5e-166;  
 Matches 995; Conservative 0; Mismatches 166; Indels 26; Gaps 5,  
 187 ATACGACATCTGATGATCTGGGAAATTCGCTGCTGATGATGATGAT 246  
 Db 239544 AGACAGACCTCTGGTACCTGGGCAATTCCTTCCTTACCGAATGAGAGC 239485  
 Qy 247 TGGATATAGACGAGTTGATTT--ATTATGAGTAGACGCTCACTACACAT 303  
 Db 239484 TGAATGTAGACAGCATGATTTATTCATTCGTGAAGTAGAGCTCACTACAGCAGCTT 239425  
 Qy 304 CCAGGATTTAACTACTTTTTCAGCATCTCACCTGAGCTTATCAATTTGAT 363  
 Db 239424 TAACTCTCTGCA--GTGCGACAGATTTCTTCACTGTGAGCTCTTAAATTTGCTT 239366  
 Qy 364 TCTTGGG-----GGAAATATCTGGGATAGAGAGGCTATTTTAA 406  
 Db 239365 TATGGGAAAT 239306  
 Qy 407 --TAAGTAGACATCTTTTCTTCTTCTTCAAGTTGATCCAAAGAGCTGTGACATC 464  
 Db 239305 AGAAGATAGTACTTTTCTTCTTCTTCAAGTTGATCCAAAGAGCTGTGACATC 239246  
 Qy 465 CATTGATTCACCTTTAAATCAAAATAGACAGACAGAGAGAGAGCAATGCTCTGA 524  
 Db 239245 AGTTGATTCACCTTTAAATCAAAATAGACAGACAGAGAGAGAGCAATGCTCTGA 239189  
 Qy 525 GTGGAACCTGATGCTTATTTATCTCGAGAAAGAGGTCCGAGTTCCCACTCTTC 584  
 Db 239188 GTGGAGACTGTAGCGTTATTTATCTCGAGAAAGAGGTCCGAGTTCCCACTCTTC 239129  
 Qy 585 CTGAGGTGTAGAGCTGATGATGCGGGGCTCAAGTTATTTTATCTCGCAATTCACATGGA 644  
 Db 239128 CTGAAGTCACTAGAGCTGAATGTTGGGGCCAGGTTTATCTCGCAATTCACATGGA 239069  
 Qy 645 TAAGCATCCCTCATTCCTCTGAGAAATTTTCCCAAGAGAGACAGGCTAAATG 704  
 Db 239068 TAAGTATCCCATTCCTCTGAGAAATTTTCCCAAGAGAGACAGGCTAAATG 239009  
 Qy 705 ACTCTGCAAGAGCTCAAGAGAGTTTTCATTTGAGAGAGATTCCTGTTCCGT 764  
 Db 239008 ACTCTGCAAGAGCTCAAGAGAGTTTTCATTTGAGAGAGATTCCTGTTCCGT 238949  
 Qy 765 AATATCTGACATCTCAAG 824  
 Db 238948 AATATCTGACATCTCAAG 238889  
 Qy 825 GAAAGCTGAG 884  
 Db 238888 GAAAGCTGAG 238829  
 Qy 885 CCGATGAATCAAGCAAGCCCAAGATGATTTGCGACAGTGAATTTGAAGATCCCTCC 944  
 Db 238828 CTGAGAGAGTCAAGCAAGCCCAAGATGATTTGCGACAGTGAATTTGAAGATCCCTCC 238769

[illegible]

```

OY 1131 GAGAACTTGAATGAAGCAGAGACCCCTGATGAGCCCGAAGAAAGATACCTCCAGAT 1190
DB 177237 GAGAACTTGAATGAATGAAGACCCCGACCGAGCTCCAGAAAGATACCTCCAGAT 177296
OY 1191 TTTATCTCAATTCAGACCTCGAAGAGGCTTTGATATGTTGTCAAGCTGAGATTCC 1250
DB 177297 TTTATCTCAATTCAGACCTCGAAGAGGCTTTGATATGTTGTCAAGCTGAGATTCC 177356
OY 1251 ACATGTGCTGCTGATACCTGATGCTGAGAGATCTTTTCATCAACCAATATACGATGACA 1310
DB 177357 ACATGTGCTGCTGATACCTGATGCTGAGAGATCTTTTCATCAACCAATATACGATGACA 177416
OY 1311 AGATCTGCTGATACCTGATGCTGAGAGATCTTTTCATCAACCAATATACGATGAG 1351
DB 177417 AGATCTGCTGATACCTGATGCTGAGAGATCTTTTCATCAACCAATATACGATGAG 177457

RESULT 12
AC112599
LOCUS
DEFINITION Rattus norvegicus clone CH230-112A20, *** SEQUENCING IN PROGRESS ***
SESSION AC112599
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 249703)
Nuzny,D.,Marie, Metzker,M.,Lee, Adyranzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amn,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benhammed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Cesaar,H., Centar,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Diya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Dava,B., Davies,K., Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falck,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabriel,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,B., Geer,K., Gill,R., Grady,M., Guerra,T., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Huliy,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,T., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,M., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuwa,L., Louissege,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Matliffe,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Mlosoavjevic,A., Miner,G., Minda,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,N., Nguyen,N., Norris,S., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwankweli,O., Okwunonu,G., Olampongon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Prims,J., Pui,L., Puazo,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Riley,B., Rodley,T., Rojas,A., Roosa,M., Rose,R., Ruiz,S., Riggs,E., Rives,C., Rodley,T., Rojas,A., Roosa,M., Rose,R., Ruiz,S., Riggs,E., Sanders,W., Savary,G., Scherer,S., Scott,C., Shadmans,S., Shen,H., Shetty,J., Shvartsberg,A., Sison,I., Sisti,C., Shadmans,S., Shen,H., Sneed,A., Sodergren,E., Song,X.Z., Sorelle,C.D., Smales,D., Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umami,K., Vlasar,R., Verz,V., Villaseca,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,D., Warren,R., Wei,X., White,S., Williams,G., Wilson,R., Wlezyk,R., Wodden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhou,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 249703)
REFERENCE
Mortley,K.C.
Direct Submission
Submitted (22-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 249703)
REFERENCE
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21743383.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/atlantis/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GROR
Center clone name: CH230-112A20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 233268 bases at least 940
Consensus quality: 237476 bases at least 930
Estimated insert size: 261159; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
----- NOTE: This sequence may represent more than one clone.
----- NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
1 249703: config of 249703 bp in length.
location/Qualifiers
1. 249703
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-112A20"
1. 1199
/clone="wgs_end_extension"
/clone_end="17"
4045..5968
/clone="wgs_end_extension"
/clone_end="177"
complement(6997..7899)
/clone="clone_boundary"

```







```

Qy 3309 TGCATTATATAGTGTGTCGCAAGCACTCTTAATTTTATGCGTGTGTCATG 3368
Db 161091 TTT---ATCACTAGTGTGTCGCAAGCAATTCCT-ATTTTGTGTTGTGTCGTGTG 161146
Qy 3369 TGTGTATGTGTAT 3381
Db 161147 TGTGTGTGTGTGT 161159

RESULT 13
BC049734 781 bp mRNA linear ROD 01-APR-2003
LOCUS BC049734 Mus musculus, clone IMAGE:6771233, mRNA.
ACCESSION BC049734
VERSION BC049734.1 GI:29436685
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE Bukaryoca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 781)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcgdb-remail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Boedet, Yaron Butcherfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candace McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Pribhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: lnl_plate: 46 Row: 9 Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
location/Qualifiers
source 1..781
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6771233"
/tissue="Testicle, mouse"
/clone_lib="NIMH_MGC_169"
/lab_host="DH10B"
/notes="Vector: pDNR-LIB"
BASE COUNT 348 a 146 c 118 t
ORIGIN
Query Match 10.6%; Score 367.4; DB 10; Length 781;
Best Local Similarity 80.2%; Pred. No. 8.5e-69;
Matches 463; Conservative 0; Mismatches 96; Indels 18; Gaps 2;
Qy 1347 GTGAGCTTCCAGATGCTGACCTGACCTGCTGCTGCTGCAAAAGGAAAGTG 1406
Db 101 GTGAGCTTCCGAGTGTCTCTCTGATGTGACTGCTGCTGCAAGATGGCAAGGAG 160

Qy 1407 ACAAAAGAGGAGAGGAGGACCTCTTGCATGACCTCTCCACATCTAGTCGAGCC 1466
Db 161 ACN---AGGAGAGAGGAGGACCTCTGCAATGACCTCTCCACTTCAGTCGACAGCC 217
Qy 1467 AGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1526
Db 218 AGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 277
Qy 1527 ACATCGACATCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 1586
Db 278 ACATCGACATCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 337
Qy 1587 AGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1646
Db 338 AGATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 397
Qy 1647 TGAGCAGCAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1706
Db 398 TAAGCAGCAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 457
Qy 1707 GTATCGAGGATTTCTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1766
Db 458 GTATCGAGGATTTCTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 517
Qy 1767 GGCATCTGAACTTTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1826
Db 518 GGCATCTGAACTTTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 566
Qy 1827 AAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1886
Db 567 ---GTAATCGCACTTTGAAATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 622
Qy 1887 AATACACTTAATGATGACATTTCTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1923
Db 623 AAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 659

RESULT 14
EX323465 175059 bp DNA linear HTG 06-JUN-2003
LOCUS EX323465
DEFINITION Danio rerio clone DKEVP-9C6, *** SEQUENCING IN PROGRESS ***, 3
unordered pieces.
ACCESSION EX323465.4 GI:31559295
VERSION HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULFILL.
KEYWORDS Danio rerio (zebrafish)
SOURCE Danio rerio
ORGANISM Danio rerio
REFERENCE Bukaryoca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 175059)
AUTHORS McLaren, S.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
fish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Jun 9, 2003 this sequence version replaced gi:29825526.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zkp9c6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 174655 bases at least Q40
Consensus quality: 174734 bases at least Q30
Consensus quality: 174818 bases at least Q20
Insert size: 176642; 1.7% error; agarose-fp
Quality coverage: 10.25x in Q20 bases; sum-of-conigs Quality

```



\* 19034 66645: contig of 47612 bp in length  
 \* 66646 66745: gap of 100 bp  
 \* 66746 73558: contig of 6813 bp in length  
 \* 73559 73659: gap of 100 bp  
 \* 113762 113761: contig of 40103 bp in length  
 \* 113762 113862: gap of 100 bp  
 \* 113862 193754: contig of 79893 bp in length  
 \* 193755 193855: gap of 100 bp  
 \* 193855 200467: contig of 6613 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1..200467  
 source  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="CH211-119P14"  
 /clone\_1db="CHORI-211"  
 1..9514  
 /note="assembly fragment:01971"  
 fragment chain:1  
 clone end:SP6  
 vector side:left  
 9615..14582  
 /note="assembly fragment:01296"  
 fragment chain:1  
 14683..16933  
 /note="assembly fragment:00946"  
 fragment chain:1  
 19034..66645  
 /note="assembly fragment:00199"  
 fragment chain:1  
 66746..73558  
 /note="assembly fragment:02277"  
 fragment chain:2  
 73659..113761  
 /note="assembly fragment:01029"  
 fragment chain:2  
 113862..193754  
 /note="assembly fragment:00416"  
 fragment chain:2  
 193855..200467  
 /note="assembly fragment:02202"  
 fragment chain:2  
 clone end:T7  
 vector side:right

BASE COUNT 64222 a 35137 c 35671 g 64740 t 700 others  
 ORIGIN

Query Match 9.2%; Score 317.6; DB 2; Length 200467;  
 Best Local Similarity 63.5%; Pred No. 5,2e-58;  
 Matches 532; Conservative 0; Mismatches 279; Indels 27; Gaps 2;

Db 514 AATGCTCTGAGTGAACCTGATGCTTATTAATCTCTGAGAAACAAGGTCGCACTTC 573  
 147426 AATGACCTGACGAAATGACGAGACTATCAACGCCACAGACAGATGATGCTCA 147485  
 QY 574 CAATCTCTCCCTGAGGTGAGAGCTGAATGTCGGGGGTCAAGTTTATTTACTCGCCA 633  
 Db 147486 GAGTGTCTCTCTGATGTGTGAGCTCAATGTAGTGAAGAGTGTACTACATCGCCA 147545  
 QY 634 TTCCACATTGATAAGCATCCCTCATCTCCCTCTGTGAAAAATGTTTCCCAAGAGAGA 693  
 Db 147546 TGTCACTCTCACAGTGTGCAAACTCACTGCTGGTAAATGTTCTCTCTAAAAAAAGA 147605  
 QY 694 CAGGCTAATGATCTAGCCAAAGACTCCAAAGGAAGTTTTCATTGACAGAGATGAT 753  
 Db 147606 CATTTCTAACACCTCAACGAGGACATCAAGGACGCTACTCATCGACCGGACGAT 147665  
 QY 754 CTGTTCCTGATATTTCTGAGCTATCTCAGGAGCAGGAGAGTGTCTCTGCTGATCACT 813  
 Db 147666 TCTTTTAGTACGTGCTGAGCTATCTCGGATTAAGACTGTGCTCTCCGATTTAT 147725  
 QY 814 TCCAGAAAAGGAAGACTGAAGGAGAGCTGAATTAATCTCCAGTCCCAAGACTGTGCA 873

Db 147726 TCCGAGAAAGGAGAGGCTGAACCGGAAGCTGAGTTTTCACAGTCCCGAGCTGTCA 147785  
 QY 874 ACTCTGACCCCGAGTGAATCAAGCAAAAGCCAGATGAATTTGCGACAGAGATTTGA 933  
 Db 147786 AATCTT-----AACCCAGATGATTAATAGTCAAGATTTTGA 147824  
 QY 934 AATATCTCTCCCAAGGAAGCAGACAAAGATCTGCCCCCTTCTCTCTGCTCCGCGA 993  
 Db 147825 CGAAGCATCCCAAGGAAAGCAGACAGAGGTTATATCTCAGCTCTTACTTGAAGCGCGGA 147884  
 QY 994 CCGCAAGTGGGTTTCAATTAATGTGGTTACAGAGATCTCGACCTTGCGCAGAGAGG 1053  
 Db 147885 CCGAGCTACCGGCTTCAATCAAGCTCGGTGACAAAGGCTCGTGGCAATTCGGGAGGAGCA- 147943  
 QY 1054 ACAGCAGATGCCAAAGTTTCGAGAGATTCCCGGATTTTGGTTGTGAAGATTTCTT 1113  
 Db 147944 -----CTGATCCCAAGCCCGCGGAATACCCAAATCTTCAATTTGCGGAAGTGGTCT 147998  
 QY 1114 GGCATAAAGATCTTTGAGAGAACTTGAATGAAGAAGAGACCTGATCGAGCCCGAGA 1173  
 Db 147999 GCGAAGAAAGTTTTCGCGACGCACTAAGAGAGCCGGATCTGACAGACCCCGGA 148058  
 QY 1174 AAGATTAACCTCCAGATTTTATCTGAAATTCAGACACTGGAAGGCTTTGATATGTT 1233  
 Db 148059 GCGTTACACTTCTCAGTTTACTGAAAGTTTCGCCACCTGAGCAGAGCTTTGATATGCT 148118  
 QY 1234 GTCAAGTGTGATTTCAATGAGTGTGCTCTCACTCATCGGTGACAGCATTTTATCA 1293  
 Db 148119 CCGGAGAGCGGGTTCCACATCTGTGCGATTTCACTACACCAATCTCCACAA 148178  
 QY 1294 CCAATATACAGATGACAAAGATCTGTCAAGCTTACATGAATATGTCTTTACCGTAG 1351  
 Db 148179 CAGGATCTGATGATGATATGATATCTGTCCAAACACAGAGATGATGCTTCTATCTGTA 148236

Search completed: September 6, 2003, 06:58:49  
 Job time : 12244 secs



```
/ Sequence 1, Application US/09328965
/ Patent No. 6501008
/ GENERAL INFORMATION:
/ APPLICANT: Nevins, Donald J.
/ APPLICANT: Simmons, Carl
/ TITLE OF INVENTION: The Regents of the University of California
/ FILE REFERENCE: 023070-096600US
/ CURRENT APPLICATION NUMBER: US/09/328,965
/ EARLIER FILING DATE: 1999-06-09
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: PatentIn Ver. 2.1
/ LENGTH: 1091
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: maize coleoptile endo-1,3,1,4-beta glucanase cDNA
/ NAME/KEY: CDS
/ LOCATION: (68)..(979)
/ OTHER INFORMATION: endo-1,3,1,4-beta glucanase
US-09-328-965-1

Query Match          2.1%; Score 73.8; DB 4; Length 1091,
Best Local Similarity 65.5%; Pred. No. 3.2e-08;
Matches 108; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3304 TAAATTGTCATTAATAGTGTGTGCCAGCACTCTCTAATTTGTTATTCGTGTGCT 3363
DB 918 TAAATTCCTCATTTTCTTCTCTATGATTTCTTTCTTTCTTTCTTTCTTTTAT 977
QY 3364 GCATGCTGTATGTATCATCAGTAATTAAGCAATTCGATGATTAATAAAAAA 3423
DB 978 GATCGCAATTAAGTTGAGAGCGGTAAAAAATTAATAAAAAAATTAATAAAAA 1037
QY 3424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
DB 1038 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1082

RESULT 3
US-09-620-312D-706/C
Sequence 706, Application US/09620312D
Patent No. 6569662
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Abundl, Vinod
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Weinman, Tom
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Yang, Yonghong
/ APPLICANT: Zhou, Jian-Rui
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yundang
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: John Fillingbaas
/ APPLICANT: Dmanac, Radoje T.
/ TITLE OF INVENTION: No. 6569662el Nucleic Acids and
/ FILE REFERENCE: 784CIP28
/ CURRENT APPLICATION NUMBER: US/09/620,312D
/ EARLIER FILING DATE: 2000-07-19
/ PRIOR APPLICATION NUMBER: 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: 09/488,725
```

```
/ PRIOR FILING DATE: 2000-01-21
/ NUMBER OF SEQ ID NOS: 1105
/ SOFTWARE: PL_FL_genes Version 1.0
/ SEQ ID NO 706
/ LENGTH: 4055
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (2515)..(3519)
US-09-620-312D-706

Query Match          2.1%; Score 73.4; DB 4; Length 4055;
Best Local Similarity 58.4%; Pred. No. 6.9e-08;
Matches 128; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 3250 AAAGACCAAGTTTATTTTCAGCATTCCTCATGCAATTCAGTGTAAACAAAAATATT 3309
DB 500 AAAGCAAAATGTTTTCAGATTGTTATTAATAAGTTATTCATACCAATMAAA 441
QY 3310 GTCAATTATAGTTGTGTGCCAGCACTCTAATTTGTTTATTCGCTGTGCATGT 3369
DB 440 GTGTACAACAGCATTTCTGTAAATTAATTAATGTTTCAGTTGTAATTTGGTATT 381
QY 3370 GTGATGTGTATCAGATTAATAAGCAATTCGATTAATAAAAAAATTAATAAAAA 3429
DB 380 TTTCGCATGCGTTTATTAATTAATTAATGCTTTTGAATAAAAAAATTAATAAAAA 321
QY 3430 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
DB 320 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 282

RESULT 4
US-09-996-243-114
Sequence 114, Application US/09996243
Patent No. 6478825
/ GENERAL INFORMATION:
/ APPLICANT: Ashkenazi, Avi J.
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerlitsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Kijavita, Ivar J.
/ APPLICANT: Napier, Mary A.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: P2710PIC13
/ CURRENT APPLICATION NUMBER: US/09/996,243
/ EARLIER FILING DATE: 2001-11-14
/ PRIOR APPLICATION NUMBER: 60/049787
/ PRIOR FILING DATE: 1997-06-16
/ PRIOR APPLICATION NUMBER: 60/062250
/ PRIOR FILING DATE: 1997-10-17
/ PRIOR APPLICATION NUMBER: 60/065186
/ PRIOR FILING DATE: 1997-11-12
```

1 PRIOR APPLICATION NUMBER: 60/065311  
2 PRIOR FILING DATE: 1997-11-13  
3 PRIOR APPLICATION NUMBER: 60/066770  
4 PRIOR FILING DATE: 1997-11-24  
5 PRIOR APPLICATION NUMBER: 60/075945  
6 PRIOR FILING DATE: 1998-02-25  
7 PRIOR APPLICATION NUMBER: 60/078910  
8 PRIOR FILING DATE: 1998-03-20  
9 PRIOR APPLICATION NUMBER: 60/083222  
10 PRIOR FILING DATE: 1998-04-28  
11 PRIOR APPLICATION NUMBER: 60/084600  
12 PRIOR FILING DATE: 1998-05-07  
13 PRIOR APPLICATION NUMBER: 60/087106  
14 PRIOR FILING DATE: 1998-05-28  
15 PRIOR APPLICATION NUMBER: 60/087607  
16 PRIOR FILING DATE: 1998-06-02  
17 PRIOR APPLICATION NUMBER: 60/087609  
18 PRIOR FILING DATE: 1998-06-02  
19 PRIOR APPLICATION NUMBER: 60/087759  
20 PRIOR FILING DATE: 1998-06-02  
21 PRIOR APPLICATION NUMBER: 60/087827  
22 PRIOR FILING DATE: 1998-06-03  
23 PRIOR APPLICATION NUMBER: 60/088021  
24 PRIOR FILING DATE: 1998-06-04  
25 PRIOR APPLICATION NUMBER: 60/088025  
26 PRIOR FILING DATE: 1998-06-04  
27 PRIOR APPLICATION NUMBER: 60/088026  
28 PRIOR FILING DATE: 1998-06-04  
29 PRIOR APPLICATION NUMBER: 60/088028  
30 PRIOR FILING DATE: 1998-06-04  
31 PRIOR APPLICATION NUMBER: 60/088029  
32 PRIOR FILING DATE: 1998-06-04  
33 PRIOR APPLICATION NUMBER: 60/088030  
34 PRIOR FILING DATE: 1998-06-04  
35 PRIOR APPLICATION NUMBER: 60/088033  
36 PRIOR FILING DATE: 1998-06-04  
37 PRIOR APPLICATION NUMBER: 60/088326  
38 PRIOR FILING DATE: 1998-06-04  
39 PRIOR APPLICATION NUMBER: 60/088167  
40 PRIOR FILING DATE: 1998-06-05  
41 PRIOR APPLICATION NUMBER: 60/088202  
42 PRIOR FILING DATE: 1998-06-05  
43 PRIOR APPLICATION NUMBER: 60/088212  
44 PRIOR FILING DATE: 1998-06-05  
45 PRIOR APPLICATION NUMBER: 60/088217  
46 PRIOR FILING DATE: 1998-06-05  
47 PRIOR APPLICATION NUMBER: 60/088655  
48 PRIOR FILING DATE: 1998-06-09  
49 PRIOR APPLICATION NUMBER: 60/088734  
50 PRIOR FILING DATE: 1998-06-10  
51 PRIOR APPLICATION NUMBER: 60/088738  
52 PRIOR FILING DATE: 1998-06-10  
53 PRIOR APPLICATION NUMBER: 60/088742  
54 PRIOR FILING DATE: 1998-06-10  
55 PRIOR APPLICATION NUMBER: 60/088810  
56 PRIOR FILING DATE: 1998-06-10  
57 PRIOR APPLICATION NUMBER: 60/088824  
58 PRIOR FILING DATE: 1998-06-10  
59 PRIOR APPLICATION NUMBER: 60/088826  
60 PRIOR FILING DATE: 1998-06-10  
61 PRIOR APPLICATION NUMBER: 60/088858  
62 PRIOR FILING DATE: 1998-06-11  
63 PRIOR APPLICATION NUMBER: 60/088861  
64 PRIOR FILING DATE: 1998-06-11  
65 PRIOR APPLICATION NUMBER: 60/088876  
66 PRIOR FILING DATE: 1998-06-11  
67 PRIOR APPLICATION NUMBER: 60/089105  
68 PRIOR FILING DATE: 1998-06-12  
69 PRIOR APPLICATION NUMBER: 60/089440  
70 PRIOR FILING DATE: 1998-06-16  
71 PRIOR APPLICATION NUMBER: 60/089512  
72 PRIOR FILING DATE: 1998-06-16  
73 PRIOR APPLICATION NUMBER: 60/089514  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028

	Query Match	Similarity	Best Local	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
Qy	3203	CPAAGTCCCAATGCTCTTCATGATCTGTGCTCAAGAAAGTTAAAGACACAGTTT	3268	2.0%;	69.2;	1441;	143;	0;	123;	0;	0;
Db	1169	CTCATGCGATGATCAATATGGGATTTGTTATTAAGTTAAATTAATCTGTATATATGTTGTT	1228	53.8%;	4.5e-87;		143;	0;	123;	0;	0;
Qy	3261	ATTTCAGCAATTCCTATGATCATTTGACGGGTAAACCAAAAATAATTTGTCAATTAAT	3322								
Db	1229	ATGATATGATATTTGAAAAAATATATACACATATAGGAAATTAATCTGTATTAATCTCTA	1288								
Qy	3323	TGTGTCCAGCACTCCTATTTGTTTATTTGCGGTGTGTGCATGTGTGTATGTATATC	3382								
Db	1289	TGTGTGCTAAATATCATTTATATCAAGAAATATTTGCTTGGCTGATTAATGATTAAG	1346								

```

Qy      3307 TTTGCAATTAATGTTGTGTGTCACAGACTCCTAATTTGTTTATTCGCGTGTGCA 3366
Db      2054 TTGAACATGCTTAAGTATAGCTGCTAATATCTTGTCTTCATTTGCTTCAGGCTGTGA 2115
Qy      3367 TCTGTGATGTGTATGACAGTAATATTAAGCATTTGATTAATATTAATATTAATTA 3426
Db      2114 TTATATTAAGTATCTTGACCAATTAATATTAATATTAATATTAATATTAATATTA 2173
Qy      3427 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
Db      2174 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2215

```

RESULT 7  
US-09-083-587-3  
; Sequence 3, Application US/09083587  
; Patent No. 6492138  
; GENERAL INFORMATION:



APPLICANT: Schmandt, et al.  
TITLE OF INVENTION: NOVEL SHC BINDING PROTEIN  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,587  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 01017/34451  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2246 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
DESCRIPTION: /desc = "mouse PAL cDNA"  
US-09-083-587-3

Query Match 2.0%; Score 69.2; DB 4; Length 2246;  
Best Local Similarity 64.2%; Pred. No. 5.4e-07;  
Matches 104; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 3307 TTTGCAATTATAGTGTGTGCGACGACCTCTTAATTTGTTTATGCGTGTGTGCA 3366  
DB 2054 TTGACATGCTCTTAAGTATGCTGCTTAATATCTTGTCTTCAATGCGTGTGTA 2113  
QY 3367 TGTGTATGTATGATCAGGTAATAAGCAATTGATGATTTAAAAA 3426  
DB 2114 TTATATAAAGTGTACTTGACCAAAAAAAAAAAAAAAAAAAAAA 2173  
QY 3427 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
DB 2174 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2215

RESULT 8  
US-09-594-506-37  
Sequence 37, Application US/09594506  
GENERAL INFORMATION:  
APPLICANT: Famodu, Omojaye O.  
APPLICANT: Rafalski, J. Antoni  
TITLE OF INVENTION: Thiamine Biosynthetic Enzymes  
FILE REFERENCE: BB1372 US NA  
CURRENT APPLICATION NUMBER: US/09/594,506  
CURRENT FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 60/139,556  
PRIOR FILING DATE: 1999-06-16  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 37  
LENGTH: 2406  
TYPE: DNA  
ORGANISM: Trilicium aestivum  
US-09-594-506-37

Query Match 2.0%; Score 69.2; DB 4; Length 2406;  
Best Local Similarity 56.6%; Pred. No. 5.6e-07;  
Matches 128; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 3243 AAGTTAAAGACCGCTTTATTTTCAGCATCTTCACGATTCAGTGTATCCAAA 3302  
DB 2179 AACTTCATAGCCCAATATTTTGTAGGATCCATTCATCTGCAAGCATTCACA 2238  
QY 3303 ATAAATTTGCAATTAATAGTTGTGCGCAAGCATCTCTAATTTGTTTATGCGTGTG 3362  
DB 2239 TGTCTGTGTAATTTACTTTTACACCTATCTTGTACATTTCTAATAGTAAATA 2298  
QY 3363 TGCATGTGTATGTATGATCAAGGTAATTAAGCAATTGATGATTAATAAAAAA 3422  
DB 2239 TAAAGATGTAATGATGACAGAAAAA 2358  
QY 3423 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468  
DB 2359 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2404

RESULT 9  
US-09-465-558-59  
Sequence 59, Application US/09465558  
GENERAL INFORMATION:  
APPLICANT: Morakinyo, Layo O.  
APPLICANT: Orozco Jr, Emil M.  
TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES  
FILE REFERENCE: BB1322 US NA  
CURRENT APPLICATION NUMBER: US/09/465,558  
CURRENT FILING DATE: 1999-12-17  
EARLIER APPLICATION NUMBER: 60/112,734  
EARLIER FILING DATE: 1998-12-18  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 59  
LENGTH: 2202  
TYPE: DNA  
ORGANISM: Glycine max  
US-09-465-558-59

Query Match 2.0%; Score 69; DB 4; Length 2202;  
Best Local Similarity 69.9%; Pred. No. 6e-07;  
Matches 93; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3336 CTCCTAATTTGTTTATGCGTGTGCGATGTGTATGTATCAGTAAATAG 3395  
DB 2064 CTCCTGTTGTTTGGCTGTGATCTGTATGAATTAATACGTGATTAAG 2123  
QY 3396 GCAATTGATGATTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3455  
DB 2124 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2183  
QY 3456 AAAAAAAAAAAAAA 3468  
DB 2184 AAAAAAAAAAAAAA 2196

RESULT 10  
US-08-665-716-1  
Sequence 1, Application US/08665716  
Patent No. 5789222  
GENERAL INFORMATION:  
APPLICANT: KEDLI, ROSEMARIE  
APPLICANT: REGISTER, ELIZABETH A  
APPLICANT: MASUREK, FRANK S  
TITLE OF INVENTION: PSC REDUCTASE GENE FROM ZALARION  
TITLE OF INVENTION: ARBORICOLA  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MERCK & CO., INC.

```
US-09-182-816-22

Query Match      2.0%; Score 68.6; DB 3; Length 1736;
Best Local Similarity 62.6%; Pred. No. 6,8e-07;
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Oy 3298 AAAAAAATATTGTCAATTAACTGTTGGCCAGACACTCCTAATTTGTTTTATTCGCT 3357
Db 1565 AATAAATATTATTGTGATATATATATTAATTAATAAATGAATTAAGCATGATTAATAAAAA 1624
Oy 3358 GTGTGTCATGTCGATGATGATGCACAGGTAAATAAGCATGATTAATAAAAA 3417
Db 1625 CGAATGCAATTTTATTTCCAAACCTTGCAATATTAATAAAAAAAAAAAAAAAAAAAAA 1688

Oy 3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
Db 1685 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1735

RESULT 12
US-09-182-816-24/c
Sequence 24, Application US/09182816
Patent No. 6143542
GENERAL INFORMATION:
APPLICANT: Wisniewski, Nancy
APPLICANT: Silver, Gary M.
APPLICANT: Lo, Katherine C.
APPLICANT: Brandt, Kevin S.
TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
FILE REFERENCE: FC-3-C1
CURRENT APPLICATION NUMBER: US/09/182,816
CURRENT FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: 08/989,510
EARLIER FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 1736
TYPE: DNA
ORGANISM: Ctenocephalides felis

US-09-182-816-24

Query Match      2.0%; Score 68.6; DB 3; Length 1736;
Best Local Similarity 62.6%; Pred. No. 6,8e-07;
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0

Oy 3298 AAAAAAATATTGTCAATTAACTGTTGGCCAGACACTCCTAATTTGTTTTATTCGCT 3357
Db 172 AATAAATATTATTGTGATATATATTAATTAATAAATGAATTAAGCATGATTAATAAAAA 113
Oy 3358 GTGTGTCATGTCGATGATGATGCACAGGTAAATAAGCATGATTAATAAAAA 3417
Db 112 CGAATGCAATTTTATTTCCAAACCTTGCAATATTAATAAAAAAAAAAAAAAAAAAAAA 53

Oy 3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
Db 52 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2

RESULT 13
US-09-471-528-22
Sequence 22, Application US/09471528
Patent No. 615397
GENERAL INFORMATION:
APPLICANT: Wisniewski, Nancy
APPLICANT: Silver, Gary M.
APPLICANT: Lo, Katherine C.
APPLICANT: Brandt, Kevin S.
TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
FILE REFERENCE: FC-3-C1-1
CURRENT APPLICATION NUMBER: US/09/471,528
CURRENT FILING DATE: 1999-12-27
```

EARLIER APPLICATION NUMBER: 09/182,816  
 EARLIER FILING DATE: 1998-10-29  
 EARLIER APPLICATION NUMBER: 08/989,510  
 EARLIER FILING DATE: 1997-12-12  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 22  
 LENGTH: 1736  
 TYPE: DNA  
 ORGANISM: *Ctenocephalides felis*  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (159)..(1553)  
 IS-09-471-528-22

Query Match	2.0%;	Score 68.6;	DB 3;	Length 1736;
Best Local Similarity	62.6%;	Pred. No. 6.8e-07;		
Matches 107;	Conservative	0;	Mismatches 64;	Indels 0;
				Gaps 0;

```

Oy      3398  AAAAAATATTTTGTCAATTAAATAGTGTGTCGCAAGCACTCCTAATTTGTTTATTCGGT 335
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1565  AATTAATATTTGTGATTAATATAATATGTTAAATAATGAATGTAATTACTGTGAAATAAA 162
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

Db      3358  GGTGTGCATGTGTGTATGTATCAAGGTAATAAAGCAATTGGATGATTTAAAAAAA 3417
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
          1625  CGATATGATTTTATTTCAAACTGTCAATATATAAAAAATAAAAAATAAAAA 168

```

QY	Db
3418	1685
3468	1735

RESULT 14  
US-09-471-528-24/c  
; Sequence 24, Application US/09471528  
; Patent No. 6153397

1 TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
2  
3 FILE REFERENCE: PC-3-CI-1  
4  
5 CURRENT APPLICATION NUMBER: US/09/471,528  
6  
7 CURRENT FILING DATE: 1999-12-27  
8  
9 EARLIER APPLICATION NUMBER: 09/182,816  
10  
11 EARLIER FILING DATE: 1998-10-29  
12  
13 EARLIER APPLICATION NUMBER: 08/989,510  
14  
15 EARLIER FILING DATE: 1997-12-12  
16  
17 NUMBER OF SEQ ID NOS: 35  
18  
19 SOFTWARE: PatentIn Ver. 2.0

```

1  LENGTH: 1736
2  TYPE: DNA
3  ORGANISM: Ctеноcephalides felis
4  US-09-471-528-24

```

Query Match	2.0%	Score 68.6;	DB 3;	Length 1736;
Best Local Similarity	62.6%	Pred. No. 6.8e-07;		
Matches 107;	Conservative	0;	Mismatches 64;	Indels 0;
				Gaps 0

DY 3298 AAAAATATATTGTCAATTAATAGTGTGTCACAGCATCTCTAATTGGTTATTGCGT 335  
||| ||| | | | | | | | |  
Db 172 AATAAATATTATTGTGATATAATATATAATGTTAAAAATTAATGTAAATCTGTGAATATA 113

OY 3358 GNETGTCATCTGTGTATGTGTACAGGTATAAAGCAATTGCATTAATAAAAAA 341  
| | | | |  
Db 112 CGATATGATTTTATTTCAAACCTGTCAATATATAAAAAAAAAAAAAAAAAAAAAA 53

[illegible]

RESULT 15  
US-09-634-530-22  
; Sequence 22, Application US/09634530

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

PRIOR APPLICATION NUMBER: 09/471,528  
 PRIOR FILING DATE: 1999-12-27  
 PRIOR APPLICATION NUMBER: 09/182,016  
 PRIOR FILING DATE: 1998-10-29  
 PRIOR APPLICATION NUMBER: 08/969,510  
 PRIOR FILING DATE: 1997-12-12  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 22

```

; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
; US-09-634-530-22

```

```
Query Match      2.0%; Score 68.6; DB 3; Length 1736;
Best Local Similarity 62.6%; Pred. No. 6.8e-07;
Matches 107; Conservative 0; Mismatches 64; Indels 0; Gaps 0
```

```
Oy      328 AAAAAAATATTGTGCAATTATAGTGTGCACGAAGACTCCTAA TTGGTTTATTCGGT   335  
        ||| | | | | | | | | | | |  
Db     1565 AAATAATTATTTTGTAATAATATAATGTTAAAATAATATGTAATTACTGTCAAATPAA    162
```

```

QY      3358 GGTGTGCATGTGTGTAATGTATACAGCTAATAAAGCATTGGATGATTAATAAAAAA 341
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1625 CGATATGCAATTTATTTCAAACTGTCAATATATAAAAAAAAAAAAAAAAAAAAA 168

```

**D<sub>b</sub>**      3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3466  
          |||||  
          1685 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1735

Search completed: September 6, 2003, 08:48:45  
Job time : 227 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2003, 06:58:53 / Search time 836 Seconds

(without alignments)  
9541.055 Million cell updates/sec

Title: US-10-056-884-1

Perfect score: 3468

Sequence: 1 caagcactgcctcaagctg.....aaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1537136 seqs, 114988732 residues

Total number of hits satisfying chosen parameters: 3074272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database:

Published Applications NA:\*

1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/prodata/1/pubpna/PCRT\_NEW\_PUB.seq:\*

3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/prodata/1/pubpna/PCRTS\_PUBCOMB.seq:\*

7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*

10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*

11: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*

12: /cgn2\_6/prodata/1/pubpna/US09C\_PUBCOMB.seq:\*

13: /cgn2\_6/prodata/1/pubpna/US10A\_PUBCOMB.seq:\*

14: /cgn2\_6/prodata/1/pubpna/US10B\_PUBCOMB.seq:\*

15: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*

16: /cgn2\_6/prodata/1/pubpna/US10C\_PUBCOMB.seq:\*

17: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3468	100.0	3468	US-10-056-884-1	Sequence 1, Appl1
2	769	22.2	769	US-10-056-884-3	Sequence 3, Appl1
3	205	5.9	632	US-10-060-036-4467	Sequence 4467, Ap
4	201	5.8	614	US-10-060-036-564	Sequence 564, Ap
5	167	4.8	2052	US-10-080-980-1	Sequence 1, Appl1
6	104.6	3.0	688	US-10-080-980-8	Sequence 8, Appl1
7	87	2.5	249	US-09-918-995-2111	Sequence 2111, Ap
8	80	2.3	80	US-10-056-884-8	Sequence 45, Appl1
9	79	2.3	425	US-09-824-975-451	Sequence 112, App
10	79	2.3	1492	US-09-925-299-112	Sequence 112, App
11	79	2.3	1493	US-09-925-299-112	Sequence 112, App
12	79	2.3	1493	US-10-106-699-1525	Sequence 1525, Ap
13	76.6	2.2	361	US-10-198-846-8619	Sequence 8619, Ap
14	75	2.2	442	US-09-770-444-872	Sequence 872, App
15	74	2.1	664	US-09-739-254-66	Sequence 66, Appl1
16	74	2.1	664	US-09-304-615-66	Sequence 66, Appl1

17	74	2.1	664	US-10-055-098-66	Sequence 66, Appl1
18	74	2.1	664	US-10-054-988-66	Sequence 66, Appl1
19	73.8	2.1	1992	US-10-036-842-18	Sequence 18, Appl1
20	73.4	2.1	2440	US-10-106-698-2024	Sequence 2024, Ap
21	73.4	2.1	4055	US-10-037-770-706	Sequence 706, App
22	73.4	2.1	465	US-10-198-846-1482	Sequence 1482, Ap
23	73	2.1	1814	US-09-918-995-31149	Sequence 31149, A
24	73	2.1	1814	US-10-163-866-23	Sequence 23, Appl1
25	72.6	2.1	405	US-09-918-995-37095	Sequence 37095, A
26	72.2	2.1	297	US-09-764-846-95	Sequence 95, Appl1
27	72.2	2.1	297	US-09-960-352-11420	Sequence 11420, A
28	72	2.1	165	US-10-198-846-8503	Sequence 8503, Ap
29	72	2.1	611	US-09-948-820-13	Sequence 13, Appl1
30	71.6	2.1	265	US-09-764-846-24	Sequence 24, Appl1
31	71.6	2.1	265	US-10-091-483-24	Sequence 24, Appl1
32	71.4	2.1	394	US-10-198-846-3918	Sequence 3918, Ap
33	71.4	2.1	420	US-09-918-995-17811	Sequence 17811, A
34	71.4	2.1	600	US-10-198-846-8414	Sequence 8414, Ap
35	71.4	2.1	4064	US-09-873-737A-3	Sequence 3, Appl1
36	71.4	2.1	368	US-09-834-975-47	Sequence 47, Appl1
37	71.2	2.1	498	US-10-198-846-1594	Sequence 1594, Ap
38	71.2	2.1	893	US-10-198-846-7225	Sequence 7225, Ap
39	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1
40	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1
41	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1
42	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1
43	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1
44	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1
45	71.2	2.1	1672	US-10-063-547-17	Sequence 17, Appl1

## ALIGNMENTS

RESULT 1

US-10-056-884-1

Sequence 1, Application US/10056884

Publication No. US20030032786A1

GENERAL INFORMATION:

APPLICANT: Bristol-Myers Squibb Company

TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL BETA-SUB

FILE REFERENCE: D0076 NP

CURRENT APPLICATION NUMBER: US/10/056, 884

PRIOR FILING DATE: 2002-01-24

PRIOR APPLICATION NUMBER: US 60/263, 872

PRIOR FILING DATE: 2001-01-24

PRIOR FILING DATE: 2001-02-14

NUMBER OF SEQ ID NOS: 73

SOFTWARE: Patentin version 3.0

SEQ ID NO 1

LENGTH: 3468

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (515) .. (1798)

US-10-056-884-1

Query Match

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3468; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CAAGACATGCTCAAGCTGTTTCAATGCAAGAAAGTTGCGCAAAATTAATGCT	60
DB	1	CAAGACATGCTCAAGCTGTTTCAATGCAAGAAAGTTGCGCAAAATTAATGCT	60
QY	61	TTGAACATGCGCAATTTTCTCTACCGCTAGCTATCCCAAGCATCATGAAGTGA	120
DB	61	TTGAACATGCGCAATTTTCTCTACCGCTAGCTATCCCAAGCATCATGAAGTGA	120
QY	121	GATTCGACGCTGTGTATTCATGCTCAAGTTCCCAATGAACCTGACATATATGTA	180

121 GATCGACGCTCTGTATTTCACTCAAGTCCACATGAAGAACTGACAAATATGCTAA 180  
181 AAACCAATAGGACACTGTGATGATGAGGAAATTTGGCTGCTTGCATGTAGCTTATG 240  
181 AAACCAATAGGACACTGTGATGATGAGGAAATTTGGCTGCTTGCATGTAGCTTATG 240  
241 GAAGATGATGATGAGGAGTGTATTAATTTATGATGAGGAGTGTATGATGAGGAGT 300  
241 GAAGATGATGATGAGGAGTGTATTAATTTATGATGAGGAGTGTATGATGAGGAGT 300  
301 CATCGAGGCTTTAACTACTTTTTCAGATCACTTCACTGTGACCTTTATGATTTTG 360  
301 CATCGAGGCTTTAACTACTTTTTCAGATCACTTCACTGTGACCTTTATGATTTTG 360  
361 ATTTCTTGGGGGAAAAATCTGGGATTAAGAGAGTCAATTTTAAATAGTAACTCT 420  
361 ATTTCTTGGGGGAAAAATCTGGGATTAAGAGAGTCAATTTTAAATAGTAACTCT 420  
421 TTTCCCTTTCTTCAAGTTGATCCAAAGATTAAGGCTGTGACTCATTGATGATGACCTT 480  
421 TTTCCCTTTCTTCAAGTTGATCCAAAGATTAAGGCTGTGACTCATTGATGATGACCTT 480  
481 TAAATCAAAAATAGGACGACAGAAAGAGGAGCAATGGCTGATGAGAACTGTAGTG 540  
481 TAAATCAAAAATAGGACGACAGAAAGAGGAGCAATGGCTGATGAGAACTGTAGTG 540  
541 TTAATCAAAAATAGGACGACAGAAAGAGGAGCAATGGCTGATGAGAACTGTAGTG 540  
541 TTAATCAAAAATAGGACGACAGAAAGAGGAGCAATGGCTGATGAGAACTGTAGTG 540  
541 TTAATCAAAAATAGGACGACAGAAAGAGGAGCAATGGCTGATGAGAACTGTAGTG 540  
601 GAATGTGGGGGCTCAAGTTATTTTACTGCGCATTCGACATTTGATGACATCTCTATT 660  
601 GAATGTGGGGGCTCAAGTTATTTTACTGCGCATTCGACATTTGATGACATCTCTATT 660  
601 GAATGTGGGGGCTCAAGTTATTTTACTGCGCATTCGACATTTGATGACATCTCTATT 660  
601 GAATGTGGGGGCTCAAGTTATTTTACTGCGCATTCGACATTTGATGACATCTCTATT 660  
661 CCTCTGTGAAAAATGTTTTCCCAAGAGAGACGCGTAAATGATCTAGCCAAAGACT 720  
661 CCTCTGTGAAAAATGTTTTCCCAAGAGAGACGCGTAAATGATCTAGCCAAAGACT 720  
721 CAAGGAGAGTTTTCATTTGACAGAGATGATCTTGTTCGTTATTTCTGATCTATCT 780  
721 CAAGGAGAGTTTTCATTTGACAGAGATGATCTTGTTCGTTATTTCTGATCTATCT 780  
781 CAAGGAGAGTTTTCATTTGACAGAGATGATCTTGTTCGTTATTTCTGATCTATCT 840  
781 CAAGGAGAGTTTTCATTTGACAGAGATGATCTTGTTCGTTATTTCTGATCTATCT 840  
841 AGCTGAATTAATCTCCAGCTCCAGACTTGTGAAATCTCTGACCCCGATGAAATCAAGCA 900  
841 AGCTGAATTAATCTCCAGCTCCAGACTTGTGAAATCTCTGACCCCGATGAAATCAAGCA 900  
901 AAGCCGAATTAATCTCTGACAGTGAATCTTGAAGATGCTCCCAAGAGACGACAAAG 960  
901 AAGCCGAATTAATCTCTGACAGTGAATCTTGAAGATGCTCCCAAGAGACGACAAAG 960  
961 AATCTGCCCCCTTCTCTCCCTGCTCCCTGCGACCGCAAGTGGGGTTTCATTTACTGTGG 1020  
961 AATCTGCCCCCTTCTCTCCCTGCTCCCTGCGACCGCAAGTGGGGTTTCATTTACTGTGG 1020  
1021 TTAAGAGAGATCTGACCTTGTGGGAGAGAGAGGAGCAAGGAGATGCAATTTCCGAGAGT 1080  
1021 TTAAGAGAGATCTGACCTTGTGGGAGAGAGAGGAGCAAGGAGATGCAATTTCCGAGAGT 1080  
1081 TCCCGGATTTTGTGTTGTGAAAGATTTCTTGTGGCAAAAGAGTCTTTTGAAGAACTTT 1140  
1081 TCCCGGATTTTGTGTTGTGAAAGATTTCTTGTGGCAAAAGAGTCTTTTGAAGAACTTT 1140  
1141 GAATGAAG 1200  
1141 GAATGAAG 1200  
1201 ATTCAAGCACTGGAAGAGGCTTTTGTATGTGTGTGAGAGTGTGATTCACATGTGGC 1260  
1201 ATTCAAGCACTGGAAGAGGCTTTTGTATGTGTGTGAGAGTGTGATTCACATGTGGC 1260

1201 ATTCAAGCACTGGAAGAGGCTTTTGTATGTGTGTGAGAGTGTGATTCACATGTGGC 1260  
1261 CTGTAACTCATGCTGAGACAGCATCTTTTCAACCAATATATACAGATGACAGAGTCTGTC 1320  
1261 CTGTAACTCATGCTGAGACAGCATCTTTTCAACCAATATATACAGATGACAGAGTCTGTC 1320  
1321 AAGCTAAGTGAATATGCTCTTCTTACAGGAGCTTCCAGATGTGACCTTCACTGCGA 1380  
1321 AAGCTAAGTGAATATGCTCTTCTTACAGGAGCTTCCAGATGTGACCTTCACTGCGA 1380  
1381 TGGCTGCTGCAAGATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1381 TGGCTGCTGCAAGATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1441 CTCTTCACATTTAGCTGCGACAGCGAGCTGAGGCTCTCTCCAGAGAGAGGTCAT 1500  
1441 CTCTTCACATTTAGCTGCGACAGCGAGCTGAGGCTCTCTCCAGAGAGAGGTCAT 1500  
1501 CTGTGTCTCCGTGACAGCGCCAGACCAACATCCAGACTGTGACCGTCCATCAAGAGG 1560  
1501 CTGTGTCTCCGTGACAGCGCCAGACCAACATCCAGACTGTGACCGTCCATCAAGAGG 1560  
1561 CCTGTCCAGCTGATCCAGCTCAGAGATGCGGCGGAGAGAGAGAGAGAGAGAGAGAG 1620  
1561 CCTGTCCAGCTGATCCAGCTCAGAGATGCGGCGGAGAGAGAGAGAGAGAGAGAGAG 1620  
1621 GACTTGAAGCTCCAGGAGATGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1621 GACTTGAAGCTCCAGGAGATGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1681 CTGAATGAG 1740  
1681 CTGAATGAG 1740  
1741 TCGGTTTCTGAG 1800  
1741 TCGGTTTCTGAG 1800  
1801 AAGGAGAGGCTGGGGGCGGGGAAAAAAGAGATGATTTGAATTAACCTCATTA 1860  
1801 AAGGAGAGGCTGGGGGCGGGGAAAAAAGAGATGATTTGAATTAACCTCATTA 1860  
1861 AAGGAGAGGCTGGGGGCGGGGAAAAAAGAGATGATTTGAATTAACCTCATTA 1920  
1861 AAGGAGAGGCTGGGGGCGGGGAAAAAAGAGATGATTTGAATTAACCTCATTA 1920  
1921 ATAGTCATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980  
1921 ATAGTCATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980  
1981 GGGTGAATTTCTTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
1981 GGGTGAATTTCTTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
2041 TACTGTGCTCCATGCTGCTAATCTTATTAATTAATTAATTAATTAATTAATTAATTA 2100  
2041 TACTGTGCTCCATGCTGCTAATCTTATTAATTAATTAATTAATTAATTAATTAATTA 2100  
2101 TGAAGAGCCTTGGAGATCAATTAATCCAACTGGGTTTTTCTCTCATCTTCTACCTCC 2160  
2101 TGAAGAGCCTTGGAGATCAATTAATCCAACTGGGTTTTTCTCTCATCTTCTACCTCC 2160  
2161 CTCTTGTGAATGAGGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2220  
2161 CTCTTGTGAATGAGGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2220  
2221 TTAATTTGATTTTCTTTTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
2221 TTAATTTGATTTTCTTTTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
2281 TTAATTTGATTTTCTTTTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
2281 TTAATTTGATTTTCTTTTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340

Oy	2341	CACAAGACA	CTAAGTATG	TATATACAT	GTACGCA	ACACAT	CCGAGG	CCCTAG	ACCTC	2400
Db	2341	CACAAGACA	CTAAGTATG	TATATACAT	GTACGCA	ACACAT	CCGAGG	CCCTAG	ACCTC	2400
Oy	2401	CCAAAGGCT	GTGTCTCG	CTCCGAC	AGCCCTCT	CTGTAA	TATTTG	TGAGT	AGT	2460
Db	2401	CCAAAGGCT	GTGTCTCG	CTCCGAC	AGCCCTCT	CTGTAA	TATTTG	TGAGT	AGT	2460
Oy	2461	TCGTACT	CTTCTT	AAATATCT	TTTGGG	AGATTTCC	AGCCTT	CTTCA	CAACTTC	2520
Db	2461	TCGTACT	CTTCTT	AAATATCT	TTTGGG	AGATTTCC	AGCCTT	CTTCA	CAACTTC	2520
Oy	2521	TAAATCA	AAATGACT	CTCATCA	CAAAATTTG	TATCTT	ATTGT	GAATTA	TACCT	2580
Db	2521	TAAATCA	AAATGACT	CTCATCA	CAAAATTTG	TATCTT	ATTGT	GAATTA	TACCT	2580
Oy	2581	CAGGCTCA	TTTATCT	GTCTTGT	CTGTGATTA	GAGAGAT	GAGAGAT	GAGT	GT	2640
Db	2581	CAGGCTCA	TTTATCT	GTCTTGT	CTGTGATTA	GAGAGAT	GAGAGAT	GAGT	GT	2640
Oy	2641	CAAAATCT	CTGTGT	TAATAAAAT	CAAAATCA	ATCATCA	CAAAATTTCT	GTCTAA	GT	2700
Db	2641	CAAAATCT	CTGTGT	TAATAAAAT	CAAAATCA	ATCATCA	CAAAATTTCT	GTCTAA	GT	2700
Oy	2701	ACTCA	CACTGAC	CTTCTTA	ACCTG	GAATTA	TATCA	CCCTTTCT	GCATGTTTCA	2760
Db	2701	ACTCA	CACTGAC	CTTCTTA	ACCTG	GAATTA	TATCA	CCCTTTCT	GCATGTTTCA	2760
Oy	2761	GAGTTCTTA	CTGCCCA	CAATTA	TGTGTG	GCCTTCC	CAATTA	TCCACATTA	TAAGTTCT	2820
Db	2761	GAGTTCTTA	CTGCCCA	CAATTA	TGTGTG	GCCTTCC	CAATTA	TCCACATTA	TAAGTTCT	2820
Oy	2821	GTGTTC	CTGTGT	TGTGTG	GAATTA	GAGAGACA	CAAGTA	CTTGAATTA	AGGTCGGGC	2880
Db	2821	GTGTTC	CTGTGT	TGTGTG	GAATTA	GAGAGACA	CAAGTA	CTTGAATTA	AGGTCGGGC	2880
Oy	2881	TTTTGT	TTGTTT	TAGAGAA	GTGTAT	TCCACA	CAACCTTA	TAATTTCT	TTATTA	2940
Db	2881	TTTTGT	TTGTTT	TAGAGAA	GTGTAT	TCCACA	CAACCTTA	TAATTTCT	TTATTA	2940
Oy	2941	TTTAA	CTACAA	AGCTACA	TTTACT	GTGCTG	TACGCTTTT	GTGTG	CTTGGGAT	3000
Db	2941	TTTAA	CTACAA	AGCTACA	TTTACT	GTGCTG	TACGCTTTT	GTGTG	CTTGGGAT	3000
Oy	3001	CGGGCT	TTGGCT	GTGCCCA	TCGTA	AGATTTAG	CTGTGTCA	TTTATGA	TGTCTGTACA	3060
Db	3001	CGGGCT	TTGGCT	GTGCCCA	TCGTA	AGATTTAG	CTGTGTCA	TTTATGA	TGTCTGTACA	3060
Oy	3061	ACCCAA	CAAGGTAA	CTGAAG	CTCGA	GTTAAG	GTTCAG	ATTTCTAA	TGAAC	3120
Db	3061	ACCCAA	CAAGGTAA	CTGAAG	CTCGA	GTTAAG	GTTCAG	ATTTCTAA	TGAAC	3120
Oy	3121	TTTTCA	TATTA	CTGACT	GTATAG	CAAGCA	AAAGAA	CTGTTA	TAGC	3180
Db	3121	TTTTCA	TATTA	CTGACT	GTATAG	CAAGCA	AAAGAA	CTGTTA	TAGC	3180
Oy	3181	GTCA	TGTATCTGT	ATTTTAT	CTAAG	GTACAT	TAGCTTTT	CACTA	CTGTG	3240
Db	3181	GTCA	TGTATCTGT	ATTTTAT	CTAAG	GTACAT	TAGCTTTT	CACTA	CTGTG	3240
Oy	3241	AGAA	GGTTAA	AGACAG	TTTATTTT	CAAGAT	TTCCTCA	NTGAT	TTCAG	3300
Db	3241	AGAA	GGTTAA	AGACAG	TTTATTTT	CAAGAT	TTCCTCA	NTGAT	TTCAG	3300
Oy	3301	AAATAT	TTTGTCA	TATTA	TGTTGT	GTGCCA	GACTCTTA	TTTGT	TTTGT	3360
Db	3301	AAATAT	TTTGTCA	TATTA	TGTTGT	GTGCCA	GACTCTTA	TTTGT	TTTGT	3360
Oy	3361	TGTG	ATGTGT	TATGTGT	ATCA	CGATTA	TAAAGCA	AT	TGAT	3420
Db	3361	TGTG	ATGTGT	TATGTGT	ATCA	CGATTA	TAAAGCA	AT	TGAT	3420

```

RESULT 2
US-10-056-884-3
/ Sequence 3, Application US/10056884
/ Publication No. US2003032786A1
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL BETA-SUB
/ TITLE OF INVENTION: K-beta1m2
/ FILE REFERENCE: D0076 NP
/ CURRENT APPLICATION NUMBER: US/10/056,884
/ CURRENT FILING DATE: 2002-01-24
/ PRIOR APPLICATION NUMBER: US 60/263,872
/ PRIOR FILING DATE: 2001-01-24
/ PRIOR APPLICATION NUMBER: US 60/269,794
/ PRIOR FILING DATE: 2001-02-14
/ NUMBER OF SEQ ID NOS: 73
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 3
/ LENGTH: 769
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-056-884-3

Query Match      22.2%; Score 769; DB 14; Length 769;
Best Local Similarity 100.0%; Pred. No. 2,1e-172;
Matches 769; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      393 AGGCTATTTTAAATAGTAGAGCATCCTTTCCCTTCTTCAAGTGTATCCAAAGGATA 452
DB      1  AGGTCACTTTTAAATAGTTAGCTTTCCTTTCCCTTCTTCAAGTGTATCCAAAGGATA 60

OY      453 AGGCTGTGATCCATGATGATGATGACCTTTAAATCAAAATATGACAGACAGAGAAGAGGA 512
DB      61  AGGCTGTGATCCATGATGATGATGACCTTTAAATCAAAATATGACAGACAGAGAAGAGGA 120

OY      513 CAATGGCTCTGAGTGAAGAACTGTATGCTTATTATTCCTCGAAGCAAGGGTCCGAGTTC 572
DB      121  CAATGGCTCTGAGTGAAGAACTGTATGCTTATTATTCCTCGAAGCAAGGGTCCGAGTTC 180

OY      573 CCAACTCCTCCCTGAGGTGTGAGAGCTGAATGTGCGGGGTCAAGTTATTTTATCTGCC 632
DB      181  CCAACTCCTCCCTGAGGTGTGAGAGCTGAATGTGCGGGGTCAAGTTATTTTATCTGCC 240

OY      633 ATTCCACATGTGATAGCATCCCTCATTCCTCCTGTGGAATAATGTTTTCCCAAGAGAG 692
DB      241  ATTCCACATGTGATAGCATCCCTCATTCCTCCTGTGGAATAATGTTTTCCCAAGAGAG 300

OY      693 ACAGGCTATATATCTAGACCAAGGACTCCAAAGGAAAGTTTTTCAATTGACAGATGAT 752
DB      301  ACAGGCTATATATCTAGACCAAGGACTCCAAAGGAAAGTTTTTCAATTGACAGATGAT 360

OY      753 TCTGTTCGGTAAATTCGTGACTATCTCAGGAGACAGGAGAGTGTCTGTGATCACT 812
DB      361  TCTGTTCGGTAAATTCGTGACTATCTCAGGAGACAGGAGAGTGTCTGTGATCACT 420

OY      813 TTCCAGAAAAGGAAGACTGAAAAGGGAAGCTGAATATCTTCAGTCCCAAGCTTGATCA 872
DB      421  TTCCAGAAAAGGAAGACTGAAAAGGGAAGCTGAATATCTTCAGTCCCAAGCTTGATCA 480

OY      873 AACTCTGACCCCGATGAATCAAGCAAGCCAGATGAATTCGACACAGTCACTTG 932
DB      481  AACTCTGACCCCGATGAATCAAGCAAGCCAGATGAATTCGACACAGTCACTTG 540

OY      933 AAGATGCTCTCCAGAGGAAGGACACAAGATCTGCCCTTCTCTCCCTGCTCCGCGG 992
DB      541  AAGATGCTCTCCAGAGGAAGGACACAAGATCTGCCCTTCTCTCCCTGCTCCGCGG 600

OY      993 ACCGCAAGTGGGGTTTCACTGTGGGGTTACAGAGATCTCGACCTTGGGCAAGAG 1052

```

```
Db 601 ACCCGAGTGGGTTTCATTCTGCGGTTACAGAGATCTTCCACTTTGGGAGAGAG 660
Qy 1053 GACAGCGAGATGCGCAAGTTTCGAGAGTTCCCGGATTTTGGTTTGGAGGATTTCT 1112
Db 661 GACAGCGAGATGCGCAAGTTTCGAGAGTTCCCGGATTTTGGTTTGGAGGATTTCT 720
Qy 1113 TGGCAAAAGAGTCTTTGGAGAACTTTGAATGAAGCAGACGCTTGA 1161
Db 721 TGGCAAAAGAGTCTTTGGAGAACTTTGAATGAAGCAGACGCTTGA 769

RESULT 3
US-10-060-036-4467
; Sequence 4467, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060.036
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4467
; LENGTH: 632
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 552..569
; OTHER INFORMATION: n = A,T,C or G
US-10-060-036-4467

Query Match 5.8%; Score 205; DB 14; Length 632;
Best Local Similarity 93.4%; Pred. No. 2.2e-38;
Matches 214; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 3206 AGTACCAATAGCTCTTCTATGACTTGTGCTACAGAGGTTAAAGACAGCTTTATT 3265
; 15 AGTACCAATAGCTCTTCTATGACTTGTGCTACAGAGGTTAAAGACAGCTTTATT 74
Qy 3266 TTCAGCATTCCTCATGATTCAGTGTACCAAAAATATTGTCAATTATAGTTGT 3325
Db 75 TTCAGCATTCCTCATGATTCAGTGTACCAAAAATATTGTCAATTATAGTTGT 134
Qy 3326 GTCCCAAGCATCTCTATTTGTTTATTCGCGTGTGTGATGTGTATGATCACA 3385
Db 135 GTCCCAAGCATCTCTATTTGTTTATTCGCGTGTGTGATGTGTATGATCACA 194
Qy 3386 GGTAATAAGGCAATGGATGATTAATAAAAAAAAAAAAAAAAAAAAA 3434
Db 195 GGTAATAAGGCAATGGATGATTAATAAAAAAAAAAAAAAAAAAAAA 243

RESULT 4
US-10-060-036-564
; Sequence 564, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
US-10-060-036-564

; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060.036
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 564
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 534..551, 575, 576
; OTHER INFORMATION: n = A,T,C or G
US-10-060-036-564

Query Match 5.8%; Score 201; DB 14; Length 614;
Best Local Similarity 93.3%; Pred. No. 1.9e-37;
Matches 210; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 3210 ACCAATAGCTCTTCTATGACTTGTGCTACAGAGGTTAAAGACAGCTTTATTTC 3269
Db 1 ACCAATAGCTCTTCTATGACTTGTGCTACAGAGGTTAAAGACAGCTTTATTTC 60
Qy 3270 GCATTCCTCATGATTCAGTGTACCAAAAATATTGTCAATTATAGTTGTGC 3329
Db 61 GCATTCCTCATGATTCAGTGTACCAAAAATATTGTCAATTATAGTTGTGC 120
Qy 3330 CAAGCATCTCTAATTTGTTTATTCGCGTGTGTGATGTGTATGATCAGGTA 3389
Db 121 CAAGCATCTCTAATTTGTTTATTCGCGTGTGTGATGTGTATGATCAGGTA 180
Qy 3390 ATAAAGCAATGGATGATTAATAAAAAAAAAAAAAAAAAAAAA 3434
Db 181 ATAAAGCAATGGATGATTAATAAAAAAAAAAAAAAAAAAAAA 225

RESULT 5
US-10-080-980-1
; Sequence 1, Application US/10080980
; Publication No. US20030036115A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN PORRASSIUM CHANNEL BETA-SUB
; FILE REFERENCE: D0121 NP
; CURRENT APPLICATION NUMBER: US/10/080.980
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270.132
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/278.953
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (121)..(1095)
US-10-080-980-1

Query Match 4.8%; Score 167; DB 14; Length 2052;
Best Local Similarity 64.6%; Pred. No. 4.8e-29;
Matches 267; Conservative 0; Mismatches 140; Indels 6; Gaps 1;

Qy 967 CCCCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1026
Db 705 CAGCGCGTCCAGTGTGAGCGGCGCGCGCTCGGCTGATCATCATCGGCTACG 764
Qy 1027 AGGATCTGCACTTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1086
```



Db 765 CGGCTCTACACCAATCGGCGGAGCGCGAGCGGAGCCAGATTCGGCGAGTGCGCG 824  
Qy 1087 GATTTTGGTTTGGAGAGATTTCTTGGCAAAAGAGCTTTTGGAGAACTTTGATGA 1146  
Db 825 CATACACGTTTGGAGAGAGCTGCTGCGCAAGAGAGTGTGGGAGACCTTAAAGGA 884  
Qy 1147 AAGCAGAGACCTGTAGAGCCCGCAGAAAGATACACTCCAGATTTTATCTCAATTCGA 1206  
Db 885 AAGCCGGAGACCCGAGCTCCCGGAGGCTACACCTCGCGCATTTACTCAAGTTCAA 944  
Qy 1207 GCACCTGGAAGGCTTTTATATGTGTGACAGAGTGGAGATTCACATAGTGGCTGTAA 1266  
Db 945 CTTCCTGAGAGAGGCTTCGACAAAGCTGTCCAGAGTCCGAGCTTCCAGATGTGGCTGAG 1004  
Qy 1267 CTGATCGGTGACAGATCTTT-----CATCAACCATATATACAGATGACAAATCTGATC 1320  
Db 1005 CTCACGGGACACTGCGCTTTTCCAGAGCAAGCAACAGAGAGAGACAAATCTGAGAC 1064  
Qy 1321 AAGCTACACTGAATATATGTCTTCTACCGTGAAGCTTTCAGATGTCACCTTCAC 1373  
Db 1065 CAGCTACACGAGATACGTCTTCTGACAGGAGTGAAGCTCCCGACCCCTCGC 1117

SUBJ 6  
US-10-080-980-8  
Sequence 8, Application US/10080980  
Publication No. US20030036115A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL BETA-SUB  
FILE REFERENCE: D0121 NP  
CURRENT FILING DATE: 2002-02-21  
CURRENT APPLICATION NUMBER: US/10/080,980  
PRIOR APPLICATION NUMBER: US 60/270,132  
PRIOR FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: US 60/278,953  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8  
LENGTH: 688  
TYPE: DNA  
ORGANISM: homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: wherein "N" is equal to "A", "C", "G" or "T".  
US-10-080-980-8

Query Match 3.0%; Score 104.6; DB 14; Length 688;  
Best Local Similarity 51.1%; Pred. No. 1,7e-14;  
Matches 192; Conservative 0; Mismatches 178; Indels 6; Gaps 1;

Db 563 TCGGCAGTTCCCAACTCTTCTCTGAGGTGTAAGCTGAATGTGCGGGGTCAAGTTAT 622  
Qy 186 TCCGCGAGAGCCAGCGCTCTTCCCGACATCGTGAAGCTGAACGTGGGGGCCAGAGTATC 245  
Db 623 TTTACTCGCATTCACATGTATAGCATCCCTCAATTCCTCTGTGAGAAATGTTTTC 682  
Qy 246 GTGACCCCGGCTGCTGAGGTGTGCTGCGGAGCTGCTGCTGCGGAGATGTTTCAG 305  
Db 683 CCAAAGAGAGACGGGTATATGATCTAGCAGAGACTCAAGAGGAGTGTTCATTGAC 742  
Qy 306 CAGCAGCA-----GCCGCAAGAGCTGGCCCGGAGCACAAAGCCGCTTCTTCTGAGC 359  
Db 743 AGAGATGATCTTGTTCGTTATTTATTTGAGCTATCTCAGAGACAGGAGGAGTCTCG 802  
Qy 360 CGGAGAGGCTTCTCTTCCCTTCACTCTGATTCCTGCGGAGCTTGCAGCTGCTG 419  
Db 803 CCTGATCACTTTCAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 862  
Qy 420 CCGGATTAATTCCTCGAGCGGAGCGGCTGACCGCGAGGCGGAGTACTTTCAGCTGCA 479

Qy 863 GACTGTGTCACCACTCTGACCCCGAGTGAATATCAAGCAAAAGCCAGATGATTTCTGCCAC 922  
Db 480 GAGCTCTCTNN 539  
Qy 923 AGTGACTTTGAGATG 938  
Db 540 NNNNNNNNTGACAAAG 555

RESULT 7  
US-09-918-995-2311  
Sequence 2311, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: MySeq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2311  
LENGTH: 249  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-995-2311

Query Match 2.5%; Score 87; DB 11; Length 249;  
Best Local Similarity 65.2%; Pred. No. 1.4e-10;  
Matches 161; Conservative 0; Mismatches 80; Indels 6; Gaps 2;

Qy 1236 CAGAGTGTGATTTCCACATGCTGCTGTATCATTCAGTGTGACAGATCTTTATCAAC 1295  
Db 1 CGAGGCGGCTTCCACATGAGGTGGGTAACTCTGGGACCGCCGCTTGCTGCAAC 60  
Qy 1296 AATATACAGATGACAAAGATCTGTCTAAGTCACTGAATATGTCTTACCGTGAGCTT 1354  
Db 61 AGTACCGGAGCAAGAGATCTGTGACAGCTACACGAGTAACTTTCTTCCGACCACTTC 120  
Qy 1355 --TCCAGATGTCACCTTCACACTGCGATTGCTGTGCAAGATGCAAG--GTGACA 1409  
Db 121 AGAAATATGATATCACTTAACAGACATGATAGATGATACCAAGTCACTGATA 180  
Qy 1410 AAGAAGGAGAGAGGAGAGTGTGCAATGACCTTCCACATCTAGCTGAGAGGCAAGT 1469  
Db 181 AAGAGATGAAATGGGACTTCTGGAATAGCTCTTCACTTCCAGTTGGAGACGCAATT 240  
Qy 1470 CTGAGGC 1476  
Db 241 CAGAGGC 247

RESULT 8  
US-10-056-884-8/c  
Sequence 8, Application US/10056884  
Publication No. US20030032786A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL HUMAN POTASSIUM CHANNEL BETA-SUB  
FILE REFERENCE: D0076 NP  
CURRENT APPLICATION NUMBER: US/10/056,884  
CURRENT FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: US 60/263,872  
PRIOR FILING DATE: 2001-01-24  
PRIOR APPLICATION NUMBER: US 60/269,794  
PRIOR FILING DATE: 2001-02-14  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8

LENGTH: 80  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Oligonucleotide Modified To Contain Biotin at the 5 P  
US-10-056-884-8

Query Match 2.3%; Score 80; DB 14; Length 80;  
Best Local Similarity 100.0%; Pred. No. 3.1e-09;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 783 GGACAGCAGCAGTGTCTGCTGCTGATCCTTCCAGAAAAAGAGCTGAAGGAGAG 842  
DB 80 GGACAGCAGCAGTGTCTGCTGCTGATCCTTCCAGAAAAAGAGCTGAAGGAGAG 21

843 CTGAATCTCTCCAGCTCCCA 862  
CTGAATCTCTCCAGCTCCCA 1

RESULT 9  
US-09-834-975-451/c  
Sequence 451, Application US/09834975  
Patent No. US20020110815A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Brown, Jeffrey  
APPLICANT: Bolt, Andrew  
APPLICANT: Van Hufel, Christophe  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS  
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
TITLE OF INVENTION: OF HUMAN CANCERS  
FILE REFERENCE: MRI-016B  
CURRENT APPLICATION NUMBER: US/09/834,975  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/197,538  
PRIOR FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 1046  
SOFTWARE: PaetSeq for Windows Version 4.0  
SEQ ID NO 451  
LENGTH: 425  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(425)  
OTHER INFORMATION: n = A,T,C or G  
US-09-834-975-451

Query Match 2.3%; Score 79; DB 10; Length 425;  
Best Local Similarity 50.8%; Pred. No. 1.5e-08;  
Matches 184; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 3107 AATGAACTACTCTTTTCAATACCTGATGTATGACAGCGCAAAAGAACT 3166  
DB 381 AAAAAAAAAAATATTTTTTTTTTTTTTTTTTTTCCAAAAAATTTTTTTTCC 322  
QY 3167 GTTAATAGCAATCGTCATGTAATCTGTATTTTACAAAGTAAAGCTTTTCA 3226  
DB 321 CTTTTTTTTTTTAAAAAGTTTTTTTTTTTAAAAAATTTTTTTTAAAAAATTT 262  
QY 3227 AGACTTGCTCAAGAGGTTAAAGACAGTTTATTTTTCAGATCTCTGATCAT 3286  
DB 261 GGGGGGATTTTTTTTTTTTTTTTGGAAAAATTTTTTTTTTTTTTTTTTTTAA 202  
QY 3287 CAGTGTAAACAAAAAATTTTGTCAATTAATGTTGTGTGCAAGCACTCTTA 3346  
DB 201 GGGGGGGGAAAAAATTTTGTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 142  
QY 3347 TTTTATGCGTGTGTGTCATGTGTATGTATCAAGGTAAATTAAGCAATGTG 3406  
DB 141 TTTTATGCGTGTGTGTCATGTGTATGTATCAAGGTAAATTAAGCAATGTG 82

QY 3407 AAA 3466  
DB 81 AAA 22  
QY 3467 AA 3468  
DB 21 AA 20

RESULT 10  
US-09-925-299-112  
Sequence 112, Application US/09925299  
Patent No. US20020055627A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 112  
LENGTH: 1492  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (8)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1487)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1491)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-299-112

Query Match 2.3%; Score 79; DB 9; Length 1492;  
Best Local Similarity 70.2%; Pred. No. 3.3e-08;  
Matches 106; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 3318 ATAGTGTGTGCGCAAGCACTCTTAATTTGTTTATGCGTGTGTGATGTATGT 3377  
DB 1276 AGAATATATATGAGGCAAGTCAATGATGATGATGATGATGATGATGATG 1335  
QY 3378 GTATCAGGTATTAAGCAATGATGATGATGATGATGATGATGATGATGATG 3437  
DB 1336 AATATATTTTATTAAGAAATTTGATTTATCAAAAAAAAAAAAAAAAAAAAA 1395  
QY 3438 AAA 3468  
DB 1396 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1426

RESULT 11  
US-09-925-299-112  
Sequence 112, Application US/09925299  
Patent No. US2003040617A9  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12



```

/ APPLICANT: Hamilton, Carol M.
/ APPLICANT: Price, Jennifer L.
/ APPLICANT: Raines, Tracy M.
/ APPLICANT: Yu, Yang
/ APPLICANT: Rameaka, Joshua G.
/ APPLICANT: Page, Amy
/ APPLICANT: Matthew, Abraham V.
/ APPLICANT: Ledford, Brooke L.
/ APPLICANT: Moessner, Jeffrey P.
/ APPLICANT: Haas, William David
/ APPLICANT: Garcia, Carlos A.
/ APPLICANT: Kricker, Maja
/ APPLICANT: Slader, Ted
/ APPLICANT: Davis, Keith R.
/ APPLICANT: Allen, Keith
/ APPLICANT: Holman, Neil
/ APPLICANT: Hurban, Patrick
/ TITLE OF INVENTION: Expressed Sequences of Arabidopsis
/ FILE REFERENCE: 2027 (PAPA-016PRV)
/ CURRENT FILING DATE: 2001-01-26
/ PRIOR FILING DATE: 2000-01-27
/ NUMBER OF SEQ ID NOS: 999
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO: 872
/ LENGTH: 442
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-09-770-444-872

```

```

Query Match      2.2%; Score 75; DB 9; Length 442;
Best Local Similarity 64.9%; Pred. No. 1.4e-07;
Matches 111; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 3298 AAAAAAAAAATTTGTCATTAATAGTGTGTCACAGCACTCCTAATTTGTTATTCGCT 3357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 AAAAAAGGTTGGTATTAATGTTGTCATTCATTTCTATTGTTGTTCTCCG 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3358 GTGTGTCATGCTGTGTATGTATCAAGGTAAVAAAGCAATTGATGATTAATAAAAA 3417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 TTAATGTAATGAAGAGTTGTTAGCATATTAACCAAGTTCAAGTTCTTAATAAAAA 53
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3418 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    53 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2

```

```

RESULT 15
US-09-739-254-66
/ Sequence 66, Application US/09739254
/ Patent No. US20010021700A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: 49 Human Secreted Proteins
/ FILE REFERENCE: P2032P1
/ CURRENT APPLICATION NUMBER: US/09/739,254
/ CURRENT FILING DATE: 2000-12-19
/ EARLIER APPLICATION NUMBER: 09/511,554
/ EARLIER FILING DATE: 2000-02-23
/ EARLIER APPLICATION NUMBER: PCT/US99/19330
/ EARLIER FILING DATE: 1999-08-24
/ EARLIER APPLICATION NUMBER: 60/097,917
/ EARLIER FILING DATE: 1998-08-25
/ EARLIER APPLICATION NUMBER: 60/098,634
/ EARLIER FILING DATE: 1998-08-31
/ NUMBER OF SEQ ID NOS: 170
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO: 66
/ LENGTH: 664
/ TYPE: DNA
/ ORGANISM: Homo sapiens

```

```

/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (31)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (63)
/ OTHER INFORMATION: n equals a,t,g, or c
US-09-739-254-66

```

```

Query Match      2.1%; Score 74; DB 9; Length 664;
Best Local Similarity 64.0%; Pred. No. 3.1e-07;
Matches 110; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

Qy 3297 CAAAAAATTTGTCATTAATAGTGTGTCACAGCACTCCTAATTTGTTATTCGCG 3356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 CATAGCTTAAATTTATTAATTTGAGGTTTGTCTTTTCTTTTCTTTTCTTTT 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3357 TGTGTGTCATGCTGTGTATGTATCAAGGTAAVAAAGCAATTGATGATTAATAAAAA 3416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 GTATATTGCTGTATCTACTTAACCTTCAGAAATTAACGTTATATRGAAATTAATA 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 3417 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: September 6, 2003, 11:49:24  
Job time : 839 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 5, 2003, 19:09:56 / Search time 6350 Seconds

(without alignments)  
13273.689 Million cell updates/sec

Title: US-10-056-884-1

Sequence: 1 caagcaccgctgaagctc.....aaaaaaaaaaaaaaaa 3468

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

EST:

1: em\_estba:\*  
2: em\_estbu:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estp1:\*  
7: em\_estro:\*  
8: em\_estc1:\*  
9: em\_estc2:\*  
10: em\_estc3:\*  
11: em\_estc4:\*  
12: em\_estc5:\*  
13: em\_estc6:\*  
14: em\_estc7:\*  
15: em\_estc8:\*  
16: em\_estc9:\*  
17: em\_estc10:\*  
18: em\_estc11:\*  
19: em\_estc12:\*  
20: em\_estc13:\*  
21: em\_estc14:\*  
22: em\_estc15:\*  
23: em\_estc16:\*  
24: em\_estc17:\*  
25: em\_estc18:\*  
26: em\_estc19:\*  
27: em\_estc20:\*  
28: em\_estc21:\*  
29: em\_estc22:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	479	13.8	489	28	A0536411
2	442.4	12.8	810	11	AK015313
3	400.2	11.5	592	28	A0525390
4	352.2	10.2	784	14	CA463745

5	352	10.1	778	13	BU961910	BU961910 AGENCOURT
6	350.6	10.1	952	14	BY714867	BY714867 BY714867
7	338.6	9.8	489	10	BF391086	BF391086 UI-R-CA1-
8	329	9.5	2332	11	AK043351	AK043351 Mus muscu
9	327.4	9.4	2343	11	AK047519	AK047519 Mus muscu
10	324.2	9.3	2584	11	AK045439	AK045439 Mus muscu
11	322.8	9.3	2555	11	AK042569	AK042569 Mus muscu
12	310	8.9	973	13	BO713664	BO713664 AGENCOURT
13	305.4	8.8	319	9	AA320222	AA320222 EST15911
14	286.2	8.3	424	14	BY706433	BY706433 BY706433
15	286	8.2	422	11	AK006368	AK006368 Mus muscu
16	284.2	8.2	419	10	BF413617	BF413617 UI-R-CA0-
17	282.6	8.1	1694	11	AK033167	AK033167 Mus muscu
18	280.6	8.1	401	11	AK005863	AK005863 Mus muscu
19	277	8.0	943	13	BY938472	BY938472 AGENCOURT
20	270.2	7.8	400	14	BY706005	BY706005 BY706005
21	257.6	7.4	386	10	BF413618	BF413618 UI-R-CA0-
22	255.4	7.4	1710	11	AK046557	AK046557 Mus muscu
23	255.4	7.4	1898	11	AK050097	AK050097 Mus muscu
24	255.4	7.4	3552	11	AK085035	AK085035 Mus muscu
25	255.4	7.4	3572	11	AK046708	AK046708 Mus muscu
26	241.2	7.0	579	12	BU095683	BU095683 BU095683
27	241.2	7.0	584	12	BU096172	BU096172 BU096172
28	234.8	6.8	491	12	BU095666	BU095666 BU095666
29	228.4	6.6	870	14	CA470866	CA470866 AGENCOURT
30	220.2	6.3	675	10	BB653208	BB653208 BB653208
31	220	6.3	1746	11	AK082563	AK082563 Mus muscu
32	219.2	6.3	295	10	BF522617	BF522617 UI-R-GO-U
33	217.6	6.3	680	10	BB183090	BB183090 BB183090
34	217.4	6.3	324	9	AI596442	AI596442 v169b03.x
35	207.2	6.0	433	9	AA497980	AA497980 v169f03.x
36	190.6	5.8	638	12	BM947722	BM947722 UI-M-EGD
37	188.4	5.5	927	14	CA964920	CA964920 CCLX0517
38	188.4	5.4	709	13	BU294624	BU294624 BU294624
39	188.4	5.4	304	9	AV046252	AV046252 AV046252
40	186.4	5.4	594	9	AL588086	AL588086 AL588086
41	179.2	5.2	310	9	AV206002	AV206002 AV206002
42	170.8	4.9	324	9	AV206367	AV206367 AV206367
43	169	4.9	295	9	AV040180	AV040180 AV040180
44	167.6	4.8	642	12	BU031784	BU031784 BU031784
45	167	4.8	1003	29	CNS05PKF	AL348072 Tetradon

#### ALIGNMENTS

RESULT 1  
A0536411/C  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

A0536411 489 bp DNA linear GSS 18-MAY-1999  
RPCT-11-318B21.TJ RPCT-11 Homo sapiens genomic clone RPCT-11-318B21  
genomic survey sequence.  
A0536411  
A0536411.1 GI:4848101  
GSS  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 489)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
J.C.  
Use of BAC End Sequences from Library RPCT-11 for Sequence-Ready  
Map Building  
Unpublished  
Other GSSs: RPCT-11-318B21.TV  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hboettger@igmc.org  
Clones are derived from the human BAC library RPCT-11. For BAC









Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL ac:

http://image.lnl.gov

Plate: LICM1080 row: n column: 06

High quality sequence stop: 473.

## FEATURES

## source

Location/Qualifiers  
1 778  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone\_image="6742567"  
/lab\_host="DH10B (T1-phage-resistant)"  
/clone\_id="Nih\_MGC\_169"  
/note="Organ: Testicles; Vector: pDRN-LIB; Site\_1: Sfil (ggccatcggcc); Site\_2: Sfil (ggccgctcggcc); cDNA made by oligo-dt priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:  
5'-AACGACGTGATACGACGAGGTGGCCATTAGCGCGG-3' and  
5'-ATTCTAGAGCGGACGCGGCGACATG-3' (30NN-3'). Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."  
JE COUNT 226 a 180 c 199 g 160 t 13 others

## ORIGIN

Query Match 10.1%; Score 352; DB 13; Length 778;  
Best Local Similarity 85.6%; Pred No 8.9e-29;  
Matches 404; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 1347 GTGAGCTTCCGATGTCACCTCCACCTGCTGCTGCAAGAAAGGCAAGG 1406  
DB 183 GTGAGCTTCCGATGTCACCTCCACCTGCTGCTGCAAGAAAGGCAAGG 242  
QY 1407 ACMAAGAGGAGAGAGCGGACGCTTTCGATGACCTTCCACATTCAGTCCGACGCC 1466  
DB 243 ACA--AGGAGAGAGAGGCGGACCTCCGCAATGACCTGTCACCTTCAGCTGACAGCC 299  
QY 1467 AGTTCGAGGCGAGCTCTCCCGAGAGACCGTCACTCTGCTGCTGCAAGCA 1526  
DB 300 AGTCAGAGGCGGAGCTCTCCCGAGAGACCGTCACTCTGCTGCTGCAAGCA 359  
QY 1527 ACATTCGAGCTTCGAGACCGTCCATCAGAGAGGCGCTGCTGCTGCAAGCA 1586  
DB 360 ACATTCGAGCTTCGAGACCGTCCATCAGAGAGGCGCTGCTGCTGCAAGCA 419  
QY 1587 AGATTCGAGGCGGAGAGGAGCTTCTCCGATTCGCTGCTGCTGCAAGCA 1646  
DB 420 AGATTCGAGGCGGAGAGGAGCTTCTCCGATTCGCTGCTGCTGCAAGCA 479  
QY 1647 TGACGAGCAAAAAAGCTCTTAAAGAAAGCTCTTAAAGAAAGCTGAGAAAT 1706  
DB 480 TGACGAGCAAAAAAGCTCTTAAAGAAAGCTCTTAAAGAAAGCTGAGAAAT 539  
QY 1707 GTATTCAGGATTTCTTAAAAAAAATTCAGATGCTTCTGAGAGAAACATCTTT 1766  
DB 540 GTATTCAGGATTTCTTAAAAAAAATTCAGATGCTTCTTAAAGAAACATCTTT 599  
QY 1767 GGCATCTGAACCTTTTAAAGAGATATCATATTAAGAGAGGCGTGGGGCGG 1818  
DB 600 GGCATCTGAACCTTTTAAAGAGATATCATATTAAGAGAGGCGTGGGGCGG 651

RESULT\_6  
LOCUS BY714867 952 bp mRNA linear EST 17-DEC-2002  
DEFINITION BY714867 RIKEN full-length enriched, adult male testis Mus musculus  
ACCESSION BY714867  
VERSION BY714867.1 GI:27127984  
KEYWORDS EST  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

## REFERENCE

## AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 952)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Niki, K., Otsu, N., Saito, R., Suzuki, H., Yamanka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schombach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A., Oakenbush, J., Schmitt, L.M., Kanpin, A., Matsuda, H., Batilov, S., Baisel, K.W., Blake, J.A., Bratt, D., Brusc, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Fraser, K.S., Gassterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S., Gustincich, S., Hitokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lemnar, B., Lyons, E.A., Maglott, R., Maltsev, L., Marchionni, L., McKenzie, L., Miki, R., Nagashima, T., Numata, K., Ohtsuka, T., Pavan, W.J., Petosa, G., Pesole, G., Petrovsky, N., Pili, R., Pong, D.J., Reid, J., Ring, R., Ringwald, M., Sander, R., Schneider, C., Sepp, C.A., Setou, B.Z., Ringwald, M., Sander, R., Takemura, Y., Taylor, H.S., Teasdale, M., Shimada, K., Sultana, R., Wagner, U., Watanabe, S., Watanabe, Y., Wells, C., Wilming, U.G., Wymshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shizaki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koye, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shizaki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

## COMMENT

## JOURNAL

## MEDLINE

## PUBMED

## FEATURES

## source

Location/Qualifiers  
1. 952  
/organism="Mus musculus"  
/mol\_type="mRNA"

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10) 1617-1630 (2000)  
RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.



[illegible]



**TITLE**  
**JOURNAL**  
**MEDLINE**  
**FOUNDED**  
**REFERENCE**  
**AUTHORS**

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 Riken integrated sequence analysis (RISA) system: 384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 11076861

**TITLE**  
**JOURNAL**  
**MEDLINE**  
**FOUNDED**  
**REFERENCE**  
**AUTHORS**

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kaneko, S., Yamashita, I.,  
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
 Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
 Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
 Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
 Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,  
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
 Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
 Carninci, P., de Bonaldo, M.F., Brownstein, M.J., But, C.,  
 Fletcher, C., Fujita, M., Gariboldi, M., Guelincich, S., Hill, D.,  
 Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Nordone, P.,  
 Marchionni, L., Mashima, J., Mazzarelli, J., Momberegs, P., Nordone, P.,  
 Rung, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
 Sato, K., Schenbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
 Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Williams, L.,  
 Wyshak-Botke, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S.,  
 and Hayashizaki, Y.  
 Functional annotation of a full-length mouse cDNA collection  
 Nature 409 (6821), 685-690 (2001)  
 21085660  
 11217851

**TITLE**  
**JOURNAL**  
**MEDLINE**  
**FOUNDED**  
**REFERENCE**  
**AUTHORS**

The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group, Phase I & II Team.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 6 (bases 1 to 2343)

**TITLE**  
**JOURNAL**  
**MEDLINE**  
**FOUNDED**  
**REFERENCE**  
**AUTHORS**

Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagata, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayata, N., Hiramoto, K., Hirakawa, T., Hirose, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kageura, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
 Okazaki, Y., Saito, K., Saitoh, H., Sakai, K., Sakai, K., Sakazume, N.,  
 Sato, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeuchi, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 Direct Submision  
 Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsunumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)

**COMMENT**  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/  
 Location/Qualifiers

**FEATURES**  
 source

1 2343  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM,DB:B930082J01"  
 /db\_xref="Ensembl:10090"  
 /clone="B930082J01"

misc\_feature  
 1..2343  
 /note="Hypothetical protein (evidence: rcds,Protein,decoder,Longest-ORF)"

BASE COUNT 554 a 660 c 624 g 505 t  
 ORIGIN

Query Match 9.4%; Score 327.4; DB 11; Length 2343;  
 Best Local Similarity 58.4%; Pred. No. 26-26;  
 Matches 780; Conservative 0; Mismatches 456; Indels 99; Gaps 8;

560 GGGTCCGACGTTCCCACTCTTCCGAGAGGTGAGAGTGAATGTGGGGGTCAAGT 619  
 512 GGGCCCTGGGCGCCGCTGCTTCCCGAGAGTGAATGTGGGGGTCAAGT 571  
 620 TATTTCAGGCACTTCACTGATGATGATGATGATGATGATGATGATGATGAT 679  
 572 TATTTCAGGCACTTCACTGATGATGATGATGATGATGATGATGATGATGAT 631  
 680 TCCCAAG 724  
 632 TCCCAAG 691  
 725 GGAAGTTTTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 784  
 692 GCGGCTTTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 751  
 785 GACAG 844  
 752 GACAG 811  
 845 GAATTCCTCAG 904  
 812 GAATTCCTCAG 871  
 905 CC-----AGATGAATTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936  
 872 TCCCTCAG 931  
 937 -----TCCCTCAG 977  
 932 GAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 991  
 978 CCGTG-----CTCCCTGCGAGCGCAAGTGGGTTTCATTCAGTGTG 1018  
 992 GTGAGT 1051  
 1019 GGTTCAGAGAGATCTGACCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078  
 1052 GGGTACCGTGGCTCTTACACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1111  
 1079 GTTCCCGGATTTGGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138  
 1112 GTGGCGGATCATGT 1171  
 1139 TTGAATGAAG 1198  
 1172 CTTAATGAAG 1231  
 1199 AATTCAAG 1258  
 1232 AATTCAAG 1291  
 1259 GCTTGAATCATGT 1318  
 1292 GCGTGAATCATGT 1351  
 1319 TCAAGTCAAG 1375  
 1352 AGCAGTCAAG 1411



QY	1376	TGCGATTGCTGTCGAGAGATGCGAAG---GTGCAAGAGAGGAGAGCGGCACTGCT	1432
DB	1412	GAACATGAGACAGGAAACCGGACAAAGTCAACGACGACGAGAGATGAGTGGGACTTCC	1471
QY	1433	TGCAATGACCTCTCCACATCTAGCTGAGACGACGAGCTGAGGCGACCTCTCCAGAG	1492
DB	1472	TGCAATGACCTCTCCACATCTAGCTGAGACGACGAGCTGAGGCGACCTCTCCAGAG	1531
QY	1493	A-----CGTCATCTGTCGCTCCGTGACACCCGACCAATCCAGATCTGACCGCT	1546
DB	1532	AACCCAGCCACACCTGACGAGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC	1591
QY	1547	CCCATCAAGAGGCGCTGTCGACCTGATCCAAAGTGAAGATGCGCGGAGAACGAC	1606
DB	1592	CCCTCCAGAGAGGCGCTGTCGACCTGATCCAAAGTGAAGATGCGCGGAGAACGAC	1651
QY	1607	TTACTCGGATTTCTGACTTCAAGGCTCCAGGGAATGAAATGAGACGACCAAAAAAGCT	1666
QY	1652	CTCTTTCAGTCACTCATGACGACGACGACGACGACGACGACGACGACGACGACGAC	1705
QY	1667	GTTAAGAAAAAGCTCTCAATTCAGAGAGAGCTGAGAAATGATTCAGATTTCTTAA	1726
DB	1706	GTCGTGAGAGAGCTAAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1765
QY	1727	AAAAAATTCAGAGAGAGCTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1786
DB	1766	ATCCATATTCAGAGATTTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1825
QY	1787	AAGTATCATCTATTA	1801
DB	1826	AAATATGCTGTGA	1840
RESULT 10			
AK045439			
LOCUS	2584 bp	mRNA	linear
DEFINITION	Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone:B3J019K12 product:hypothetical protein, full insert sequence.		
ACCESSION	AK045439		
VERSION	AK045439.1	GI:26337364	
KEYWORDS	HTC; CAP trapper		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus		
REFERENCE			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20493574		
PubMed	11042159		
REFERENCE			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PubMed	11076861		
REFERENCE			
AUTHORS	Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamashita, I., Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gasterland, T., Gissi, C., King, B., Koehne, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pasale, G., Quackenbush, J., Schriml, L. M., Stambli, P., Suzuki, R., Tomita, M., Wagner, L., Wastio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balaguer, P., de Bonaldo, M. F., Brownstein, M. J., Bul, C., Carninci, P., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hornum, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, U., Mazzarelli, V., Mombautte, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. P., Suzuki, R., Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S., and Hayashizaki, Y.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409 (6821), 685-690 (2001)		
MEDLINE	21085660		
PubMed	11217851		
REFERENCE			
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
MEDLINE	6 (bases 1 to 2584)		
PubMed	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hyman, K., Hirooka, T., Hirozane, T., Hori, F., Imocan, K., Ishii, Y., Itoh, M., Kageura, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, N., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ono, M., Onose, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)		
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.		
FEATURES	Please visit our web site for further details.		
source	URL: http://genome.gsc.riken.go.jp/		
	URL: http://fantom.gsc.riken.go.jp/		
	Location/Qualifiers		
	1..2584		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6J"		
	/db_xref="FANTOM:DB:B3J0119K12"		
	/db_xref="taxon:10090"		
	/clone="B3J0119K12"		
	/sex="male"		
	/tissue="corpora quadrigemina"		
	/clone_lib="RIKEN full-length enriched mouse cDNA library"		
	/dev_stage="adult"		
	262..1692		
	/note="unlabeled protein product; hypothetical protein"		

CDS



(evidence: rscDS, ProCrest, decoder, Longest -ORF)  
 putative  
 /codon\_start=1  
 /protein\_id="BAC32368.1"  
 /db\_xref="GI:26337365"  
 /translation="MALKDTGSGSGSTLIPISSEWVSASSSPGAPLAAPGPCAPSPPE  
 VEINAVGOVYVTHSTLLSPDSTLAAMPSPSPGARRRQDIPRDSKARFIDSD  
 GFLFRVYLDLRDKQALPPEHPEKEREFLQTLVLLKSPKVTOKRISLANDC  
 COSLEBNVAGSGSDALLRGAAGAPEGSGAVSGVGGSGAPYKSGFTLGYG  
 SYTVRDNQADAKFRVARIWCGRIALAKEVFDLTNESRDPDQPEKXTSRFLKF  
 TYLEAFDRLEAGFPHVAVCNSSGRIAAVNOYRDKIMSYTYEIFFRPOKIVSPPO  
 EHEBRKDKVTGDSGSGSNELVTSKDSHSEASTPQNPANTQAAAADPNTLL  
 DRPSRKAPOVMPPEPKRNSLFOSLSKSRFTNLSKKVKCVKLSVEBEKKKIQDF  
 KKIHPPCFERRKROMSBLKQYGL"

BASE COUNT 681 a 636 c 646 g 621 t  
 ORIGIN  
 Query Match 938; Score 324.2; DB 11; Length 2584;  
 Best Local Similarity 58.38; Pred. No. 4,1e-26;  
 Matches 778; Conservative 0; Mismatches 458; Indels 99; Gaps 8;  
 560 GGATCGGAGTTCCTCCAACTCTCTCTCTGAGGTGATGAGTGAATGTGCGGAGTCAAGT 619  
 364 GGGCCCTGCGCCGCGTCCCTTCCCGAGGTGAGAGTGAATGTGCGGAGTCAAGT 423  
 620 TATTTTACTCGCATTCACATTGATAGCATCCTCTCTCTCTGAGGAAATGTT 679  
 424 TATGTGACCAAGCATTCGAGCTTACTCAGCGTCCGAGCAGCACTCTGCGCAGCATGTT 483  
 680 TCCCAAGAGAGAGACAGGCGTAAATGATC-----TAGCCAGAGCATCCAG 724  
 484 TACCTCTAGTCCCGGCGGCGGCGCTAGGCGCGGCGGCGCATGCGCCAGAGCAGCC 543  
 725 GGAAGTTTTCATTGACAGAGATGATCTTGTCTCTCTCTCTCTCTCTCTCTCTCT 784  
 544 GCGGCTTCTTATGACAGCGAGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 603  
 785 GACAGGAGTGTCT 844  
 604 GACAGGAGTGTCT 663  
 845 GAATCTTCCAGCTCCAGCT 904  
 664 GAGTCTTCCAGCTCCAGCT 723  
 905 CC-----AGATGAATCTGCAAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 936  
 724 TCGCTCAAGATGATCT 783  
 937 -----TGCTCCCAAG 977  
 784 GACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 843  
 978 CCCTG-----CTCCCTGCGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018  
 844 GTGAGT 903  
 1019 GGTTCAGAGATCTCTGACCTTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078  
 904 GGTACCGT 963  
 1079 GTTCCCGGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1138  
 964 GTGGGCGCATCATGT 1023  
 1139 TTGAATGAAG 1198  
 1024 CTTAATGAG 1083  
 1199 AATTCAAG 1258  
 1084 AGGTCTACTTGTGAG 1143

1259 GCCTTAATCTATGCTGACAGCATCTTCACTCAACCAATATACAGATGACAGATCTCG 1318  
 1144 GGTCGAACTCTCTGACAGCTGCGCTTTTCAACAGAGAGAGAGAGAGAGAGAGAGAG 1203  
 1319 TCAAGCTACTGATATATGCTTCTTACCGTAGAGCT---TCCAGATGTCTACCTCTAC 1375  
 1204 AACAGCTTACATGATATATATCTTCTTCCAGACCTCGAATAATGTCTACCCAGCA 1263  
 1376 TCGATATCTCTCTCAAGATGAGCAAG---GTGACAAAGAGAGAGAGAGAGAGAGAG 1432  
 1264 GAACATGAAAGACAGAAACCCGACAAAGTACAGACAGAAAGAGAGAGAGAGAGAG 1393  
 1433 TCGATATCTCTCTCAAGATGAGCAAG---GTGACAAAGAGAGAGAGAGAGAGAGAG 1492  
 1324 TCGATATCTCTCTCAAGATGAGCAAG---GTGACAAAGAGAGAGAGAGAGAGAG 1383  
 1493 A-----CGTCACTGT 1546  
 1384 AACCCAGCAAG 1443  
 1547 CCCATCAAG 1606  
 1444 CCTCAG 1503  
 1607 TTAATCCGATTTGATCTTCAAGCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666  
 1504 CTCTTCACTGATCATCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557  
 1667 GTTAAAG 1726  
 1558 GTCTGTGAG 1617  
 1727 AAAAAAATTCAGATGCTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786  
 1618 ATCCATATTCAGATGCTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677  
 1787 AAGTATCATCTATTA 1801  
 1678 AATATGAGGTGTTAA 1692

RESULT 11  
 AK042569 2555 bp mRNA linear HTC 05-DEC-2002  
 LOCUS  
 DEFINITION  
 accession  
 AK042569.1 GI:26335190  
 KEYWORDS  
 Mus musculus (house mouse)  
 HTC; CAP-trapper.  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 JOURNAL  
 Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE  
 PUBMED  
 99279253  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayashizaki, Y., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subcloning of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, Y., Ishii, K., Katsunari, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishikawa, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaetzel, K.,



Db 1024 ACTCTTAATGAGTGGACCCCTGACCGTCAAGCTTGAAGTACATCCCGCTTCTAC 1083  
 Oy 1196 CTCAAATTCAGACCTGGAAAGGCTTTTATATATGTTGTCAGAGTGTGATTCACATG 1255  
 Db 1084 CTCAGTTCACTACTTGAAGGAGGCTTGAATGATGATGCTGAGGCGGCTTCCACATG 1143  
 Oy 1256 GTGGCCTGTAACTCACTGGGTGACAGCATCTTTCATCAACCAATATACAGATGACATG 1315  
 Db 1144 GTGGCTGCACTCTCTGTCAGCTGCGGCTTTTGTCAACAGTACGAGCAAGATC 1203  
 Oy 1316 TGTCAAGCTACACTGATATATGTTCTTCTTACCGTAGCTT---TCCAGATGTCACTCA 1372  
 Db 1204 TGGAGCAGTTACACTAAATACATCTTCTTCCAGCAGCTCAGAAATATGTCACTCAAG 1263  
 Oy 1373 CACTGCAATGCTGCTGCAAGAAATGGCAAG--GTGCAAAAGAGGGAGAGCGGACG 1429  
 Db 1264 CAAGACATGAGAGAGAGAAAGCGACAAAGTCAAGACAAAGAAAGTGAAGTGGGACT 1323  
 Oy 1430 TCTTGAATGACCTCTCCACATCTATAGCTGACAGCAGCTGAGGCGAGCTTCCGAG 1489  
 Db 1324 TCTGCAATGAGCTCTCCACATCTCAGCTGTGACAGCAGCTCAGAGGCGAGCTTCCAG 1383  
 Oy 1490 GAGA-----CGGTCACTGTGTGTCCTCGTGAACAGCCAGCAACATCCAGACTGTGAC 1543  
 Db 1384 GACAACCCAGCAACCTCAGAGGCTGACACTCAGCAGCTTAACTTAACTTGGAT 1443  
 Oy 1544 CGTCCCATCAAGAAAGGCGCTGTCCAGCTGATCCAAAGTCAAGATGGCGGAGAAAGC 1603  
 Db 1444 AGACCTTCCAGAAAGCACTGTTCAATGATGTCACACACACACAGAGGAGAAAGCT 1503  
 Oy 1604 GACTTACTCCGATTTCTGACTTCCAGCTTCCAGGATATGCAATGAGCAGCAAAAAA 1663  
 Db 1504 GAACTCTTCACTGACTCATCAGCAAGTCCGAGAAACAAATCT-----CTCCAAAG 1557  
 Oy 1664 GCTGTAAAGAAAGCTCTCAATTAAGAGAGAGCTGAGAAATGTATCCAGATTTCTTA 1723  
 Db 1558 AAGCTGTGTGAGAGCTAAGTATGAGAGAGAAATGAAAGGTATTCAGATTTTAA 1617  
 Oy 1724 AAAAAAATTCAGATCGGTTTCTGAGAGAAACATCTTGGCAATCTGAATTTTA 1783  
 Db 1618 AAAATCCATTCAGATGTTGTTTCCAGAGCGCAAGCAGGCAATCTGAATCTCTC 1677  
 Oy 1784 AGGAATATCATCTATTA 1801  
 Db 1678 CAAAAATATGGTTGTA 1695

RESULT 12  
 B0713664  
 US  
 INITIATION 973 bp mRNA linear EST 16-JUL-2002  
 AGENCOURT 8480138 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6310836  
 5', mRNA sequence.  
 B0713664  
 B0713664.1 GI:21852563  
 EST.  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Tissue musculi  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.  
 1 (bases 1 to 973)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsdbs-remail.nih.gov  
 Tissue Procurement: Susan U. Sullivan, Ph.D.  
 cDNA Library Preparation: Reagen, Invitrogen Corp.  
 DNA Sequencing by: Agencourt Bioscience Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: L1M1373 row: m column: 13  
 High quality sequence start: 23  
 High quality sequence stop: 592.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..973  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6310836"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: olfactory epithelium; Vector: pCMV-Sport6.1.cdb; Site: 1; Score: 5; Note: Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Reagen, Invitrogen Corp. Note: This is a NIH\_MGC Library."  
 BASE COUNT 275 a 216 c 192 g 288 t 2 others  
 ORIGIN  
 Query Match 8.9%; Score 310; DB 13; Length 973;  
 Best Local Similarity 68.8%; Pred. No. 2,4e-24;  
 Matches 681; Conservative 0; Mismatches 242; Indels 67; Gaps 16;  
 Oy 1690 GAGAGAGCTGAGAAATGATATCAGATTTCTTA--AAAAAATTCAGATCGGTT 1747  
 Db 1 GGAAGAGCTGAGAAATGATATCAGATTTCTTGAAGATTAATTCAGATCGCTTC 60  
 Oy 1748 CCGTGAAGAAATCATCTGCGAATCTGAATCTTTAAGAGATCATCTTAAGGAGG 1807  
 Db 61 CCGTGAAGAAATCATCTTGGAGCTGATTAATTTAAGAGATCATCTTAAGGAGG 120  
 Oy 1808 GCTGGGCGGGGAAAAAAGAGCTTTTGAATTAATCAATTAAGAAAT 1867  
 Db 121 GCGTGG-----GTAGTGGCTTTGAATTAATCTCCCAAGAG 165  
 Oy 1868 TCAATTTTAAAGAAAAAATATCACTTAATGATGCACTTTCTTGAACCAATAGTTC 1927  
 Db 166 ACAATATGTTAAAGAAAAATTA-ACAATCAAGGTGCACTTTGATGATCAAT-GTCC 223  
 Oy 1928 ATTGATTAATCTAGCTACTTTAATCTTAACTTAAATCATGTAATCAACAGGTAGA 1987  
 Db 224 ATTGATTAATCTAGCTACTTTGCTTGTGCTTAACTTAACTTAACTTAACTTAA 283  
 Oy 1988 TTTCTTTGATGATGAGAGTCAAGAAATCTTTTGT--TTTGTGTTTCTT 2045  
 Db 284 TTTCTTTGATGATGAGAGTCAAGAAATCTTTTGT--TTTGTGTTTCTT 343  
 Oy 2046 CGTCCATGCTAATCTTCTT-ATATTAATGAGAGCAGCTAGTAAAGTGTGAG 2104  
 Db 344 GGTCCATGCTGAGAAATCTTAATTAATCAAGAAACAGCTAGTGTGAGTGTGAG 403  
 Oy 2105 AGGCTTGGAGATCTTTTCCCAACCTGGTTTTT----- 2141  
 Db 404 AGGCTTGGAGATCTTTTCCCAACCTGGTTTTT----- 2141  
 Oy 2142 ---CTTCATCTTCTTACTCCCTCTTGA--ATGAGGATGATGAGAAAGATCTGG 2196  
 Db 464 TTACTCTATTTTCT 523  
 Oy 2197 CCCAATGCAATAGTTTGAATTTTAAATTTTGGTTTTCT--TTGTTTATGGGTT 2253  
 Db 524 CCCGATGCTTATGTTAAGAAATTAATCTTTTCCCTTTCTTTTGTGTTTATGGGTT 583  
 Oy 2254 --GGGGGAGATGAGATTTATATGATCTTCACTCAATCTATATGTCAGATTTAT 2311  
 Db 584 GAGGGGAGATGAGAAATTTTATATGATCTTCACTCAATCTATATGTCAGATTTAT 643  
 Oy 2312 TGACTCGTATGATGATGATTTTGTG--CAACCAAGCAGCAATGATGATTAACA 2368  
 Db 644 TGACTCGTATGATGATGATTTTGTG--CAACCAAGCAGCAATGATGATTAACA 703  
 Oy 2369 CATACGACAGATGACAGGCTTAACTTCCCAAGGCTGTGCTCTCTCTCTCTCTCTCT 2428

Df	704	AAGATGACACGACCCTAAGGGC--TGACAGCAGAGGGGTAACTTACTATTCAAGCTG	761
Oy	2429	CCCTCTTTGAATATTTTCAGATGATGCTGTGATCTTCTTTTAATATCTTTGGG	2488
Df	762	CCC-CTACAGAGCATTTGAGA-CACAGAGCTCTGGCTATTATT-AAACCCCTCGGG	818
Oy	2489	AAGATTTCCAGGCTTTCTT-----CACAGACTTTTCAATCAATACATCTGCATATC	2544
Df	819	CAAATTTCCAGGCTCTCCCTGCAGACACTTTTAAAGCTGAATTAGGCCCATCATC	878
Oy	2545	AACAATGTTATTCCTTATTGGAATTAATACCTCAGCTTCATTTACTGTTGCT	2604
Df	879	AAAAATTCCTCTCTTAAATTTGAAAAAACCCCAAGGCTCTCTGAAATAATAGG	938
Oy	2605	CTTGTCTGCTAATAGAGAGATGAGAGA	2634
-	939	CCCTTCCCCACCTAAGAAGCTGGGAAA	968
RESULT 13			
LOCUS	AA132022	319 bp	mRNA linear EST 21-Apr-1997
DEFINITION	EST15911 Embryo, 8 week I Homo sapiens cDNA 5' end, mRNA sequence.		
VERSION	AA132022		
KEYWORDS	AA132022.1 GI:1984264		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 319)		
AUTHORS	Adams,M.D., Kertjave,A.R., Fleischmann,R.D., Fulton,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weisbrock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.P., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghegan,N.S., Glöckel,A., Gruber,I.N., Hume,D.A., Hunkeler,F.-S., J., Kelley,J.M., Kelley,U.C., Li,H.-I., Marraffini,S.H., Metzger,J.W., Morino-Palancues,R.F., McDonald,J.B., Nguyen,D.T., Peligro,S.M., Phillips,C.A., Ryder,S.E., Scott,J.U., Sauden,D.M., Shiller,R., Small,K.V., Sprague,T.A., Utterback,T.R., Wetland,J.B., L.Y., Bedarick,D.P., Cao,L., Cepeda,M.A., Coleman,J.B., Collins,B.J., Dime,D., Feng,D.-P., Fertile,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunchik,C., Hunglung,J., Li,H., Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,Y., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Pannon,Vener,J.C., Rosen,C.A., Haseltine,W.A., Fields,C., Frazer,C.M. and Venter,J.C.		
TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence		
JOURNAL MEDLINE PUBMED COMMENT	Nature 377 (6547 Suppl), 3-174 (1995) 756098 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlavage@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index ( <a href="http://www.tigr.org/cdb/hgi/hgi.html">http://www.tigr.org/cdb/hgi/hgi.html</a> ) Seq primer: M13 Reverse Location/Qualifiers		
FEATURES	Source		
	1..319		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="ATCC (inhouse):133650"		
	/dev_stage="Embryo, 8 wks"		
	/clone_id="Embryo, 8 week I"		
	/note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;"		

BASE COUNT	99 a	Site 1: EcoRI, Site 2: XhoI	
ORIGIN	8.8%; Score 305.4; DB 9; Length 319; Best local Similarity Pred. No. 1.4e-23; Matches 317; Conservative 0; Mismatches 1; Indels 1; Gaps 1;		
Query Match			
Db	1842	TTTGAATTAATACCTCAATGAAGATTCATTTTAAAGAAAAATTCACATATGAT	
Oy	1902	GCACTTTTGAACACATATGATTCATTCATCTCTCCCTCTTTTACTTACTGCTC	
Df	61	GCACTTTTGAACACATATGATTCATTCATCTCTCTCTCTCTCTCTCTCTCTCT	
Oy	1962	CTTACATGTAATATCCACAGGATGATTTCTTTCTAGATGTGAAGTACAGAAATCTT	
Df	121	CTTACATGTAATATCCACAGGATGATTTCTTTCTAGATGTGAAGTACAGAAATCTT	
Oy	2022	TTTTAGTTATTTGTTGTTTACTTCTGTCATGCTCTACTATCTTATATATATGAG	
Df	181	TTTTAGTTATTTGTTGTTTACTTCTGTCATGCTCTACTATCTTATATATATGAG	
Oy	2082	CCAGTACCTTAATAGTACGTGAGAGGCTGTGGAGTCAATTATCCAACTGGG-TTTT	
Df	241	CCAGTACCTTAATAGTACGTGAGAGGCTGTGGAGTCAATTATCCAACTGGGTTTTT	
Oy	2141	TCTCTCATCTCTCTCTCTC 2159	
Df	301	TCTCTCATCTCTCTCTCTC 319	
RESULT 14			
LOCUS	BY706433	424 bp	mRNA linear EST 16-DEC-2002
DEFINITION	BY706433 Riken full-length enriched, adult male testis Mus musculus cDNA clone 1700026A08 5', mRNA sequence.		
VERSION	BY706433		
KEYWORDS	BY706433.1 GI:27117598		
SOURCE	EST.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 424)		
AUTHORS	Okazaki,Y., Furuno,M., Saito,R., Suzuki,H., Yamanka,I., Kiyosawa,H., Mikado,I., Ooto,N., Kasekawa,T., Adachi,J., Bono,H., Kondo,S., Yagi,K., Tomaru,Y., Haasgawa,Y., Nogami,A., Schonbach,C., Gajobori,T., Baladrelli,R., Hill,D.P., Bull,C., Hume,D.A., Quackenbush,J., Schiml,L.M., Kanpin,A., Matsumura,H., Balov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chochoja,C., Corbett,L.E., Cousins,S., Dalla,E., Drengni,T.A., Fletcher,C.F., Forrest,J., Fraser,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimond,S., Guerinich,S., Hirokawa,N., Jackson,J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Malais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavani,W.J., Pettes,G., Pelele,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takeuchi,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wyman-Zhu,X., Zimmermann,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhou,Y., Zimmermann,M., Hayatsu,N., Hirozane-Klemliska,T., Komno,H., Nakamura,M., Sakazume,N., Seto,K., Shiraki,T., Maki,K., Kawabi,O., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Amachi,K., Ishii,Y., Itchi,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,R., Shibata,K., Shinagawa,A., Tsunishi,A., Yoshino,M., Westesson,R., Lander,E.S., Rogers,J., Birney,S. and Hayashizaki,Y.		
TITLE	Analysis of the mouse transcriptome based on functional annotation		







## SUMMARIES

9	167	4.8	2052	24	ABT09812	Human polycomb target
C 10	167	4.8	109201	24	ABO808125	Human osteoblast
C 11	114.2	3.3	854	24	ABO400554	Oligonucleotide f
C 12	114.2	3.3	854	24	ABO400554	Oligonucleotide f
C 13	109.8	3.2	1757	24	ABO133668	Oligonucleotide f
C 14	109.8	3.2	1757	24	ABO133668	Oligonucleotide f
C 15	108.6	3.1	854	24	ABO400556	Oligonucleotide f
C 16	108.6	3.1	854	24	ABO400557	Oligonucleotide f
C 17	95.6	3.0	688	24	ABT09813	K-beta M $\mu$ related
C 18	95.6	2.8	1757	24	ABO133666	Oligonucleotide f
C 19	95.8	2.8	1757	24	ABO133667	Oligonucleotide f
C 20	80	2.3	80	24	AAB46069	Antisense oligonu
C 21	79	2.3	425	22	AAS60450	Human cancer agn
C 22	79	2.3	1119	21	AAC60033	Human secreted pr
C 23	79	2.3	1492	21	AAC98102	Human colon cance
C 24	79	2.3	1493	22	AAC94433	Human colon cance
C 25	79	2.3	1493	22	ABH90031	Human polynucleot
C 26	76.8	2.2	2796	24	ABH90055	Human polynucleot
C 27	76.2	2.2	847	23	ABH06735	Drosophila melanog
C 28	75.2	2.2	2847	23	ABH06734	Drosophila melanog
C 29	75.2	2.2	1855	23	ABR43528	DNA encoding nove
C 30	75	2.2	442	24	ABP94107	Arabidopsis thali
C 31	74.2	2.1	655	32	AHC70113	Human cervical ca
C 32	74.2	2.1	887	21	AAS59297	Human secreted pr
C 33	74	2.1	664	21	AAX26336	Human secreted pr
C 34	73.8	2.1	1091	22	AAC89723	MtZ zingherin-1 g
C 35	73.8	2.1	1091	25	ABA59505	cDNA encoding mal
C 36	73.6	2.1	1392	22	AAT72748	Human prostate cat
C 37	73.6	2.1	1504	21	AAC59636	Human secreted pr
C 38	73.4	2.1	1775	23	AHA44511	Human prostate ex
C 39	73.4	2.1	2440	22	ABH43532	Human colon cance
C 40	73.4	2.1	4055	22	AAI58815	Human polynucleot
C 41	73.2	2.1	2377	21	AAC69641	Human secreted pr
C 42	73	2.1	348	22	AAAI0133	Human breast canc
C 43	73	2.1	1814	25	ABT17358	Human SLC7 relate
C 44	72.6	2.1	346	23	ABV48988	Human prostate exp
C 45	72.2	2.1	297	22	AAS29114	cDNA encoding for

## ALIGNMENTS

## RESULT 1

ID AAD46068 standard; cDNA; 3468 BP

AC AAD46068;

DT 27-DEC-20

Human K+betam2 cdna.

Human: potassium chloride

reproductive disorder:  
KW

KW spermatogenesis; reproductive

KW carcinoïd tumour; b

KW graft-versus-host di

nephrotropic; cytochrome

OS Homo sapiens.

Key	Location
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

ET / \*E

•

XX.

```
Location/Qualifiers
515..1801
/*crg= a
/product= "Human K-betam2 protein"
```





[illegible]





XX DE Human VM106R.1 homologue-encoding cDNA, SEQ ID NO:992.  
 XX XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 XX haematopoietic regulation; tissue growth; immunomodulator; activin;  
 XX inhibin; chemotaxis; chemokinesis; thrombolytic; oncogenesis;  
 XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 XX chronic inflammatory condition; proliferative retinopathy;  
 XX atherosclerosis; coronary heart disease; arterial ischaemia;  
 XX bone disorder; osteoporosis; vascular growth disorder;  
 XX tissue regeneration; wound healing; infection; immune disorder;  
 XX cell culture; drug screening; gene therapy; antiinflammatory;  
 XX antirheumatic; antiarthritic; hemostatic; antiarteriosclerotic;  
 XX cyclostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 XX antitumoral; antiviral; antitumor; ss.  
 XX Homo sapiens.  
 PN MO200157188-A2.  
 XX 09-AUG-2001.  
 PD 05-FEB-2001; 2001MO-US03800.  
 PF 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT;  
 XX WPI: 2001-457740/49.  
 DR P-PSDB; ABB11972.  
 XX Human proteins and DNA encoding sequences useful for preventing  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer.  
 PS Claim 1; Page 844-845; 1963pp; English.  
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 XX invention also relates to vectors and recombinant host cells comprising a  
 XX nucleotide of the invention, methods of producing the novel polypeptides,  
 XX antibodies against the polypeptides, methods of detecting the nucleotides  
 XX or polypeptides in a sample, and methods of identifying compounds which  
 XX bind to polypeptides of the invention. Although novel, many of the  
 XX polypeptides of the invention have homology to known proteins, thereby  
 XX giving an insight into their probable biological activities, and hence  
 XX potential therapeutic applications. The polypeptides of the invention may  
 XX have various activities, including cytokine, cell proliferation or cell  
 XX differentiation regulatory activity; tissue growth factor activity;  
 XX haematopoietic regulatory activity; stem cell growth factor activity;  
 XX immunomodulatory activity; activin- or inhibin-related activities;  
 XX chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 XX thrombolytic activities; receptor or ligand activities; or may be  
 XX involved in oncogenesis, cancer cell proliferation or metastasis.  
 XX Depending on their biological activities, polypeptides and nucleotides of  
 XX the invention are useful for preventing, treating or ameliorating medical  
 XX conditions, e.g. by protein or gene therapy. Such conditions include  
 XX cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell  
 XX disorders), chronic inflammatory conditions (e.g. asthma or arthritis),  
 XX proliferative retinopathy, atherosclerosis, coronary heart disease,  
 XX arterial ischaemia, bone disorders (e.g. osteoporosis), and abnormal  
 XX vascular growth. Polypeptides involved with tissue regeneration and  
 XX repair (or nucleic acids encoding them) may be used to promote wound  
 XX healing (e.g. of burns, incisions and ulcers), while those of viral,  
 XX immunomodulatory activities may be used in the treatment of viral,  
 XX bacterial and fungal infections in addition to immune disorders.  
 XX Polypeptides with growth factor activity may be used in cell cultures to  
 XX promote cell growth. For example, such polypeptides may be used to  
 XX manipulate stem cells in culture to give rise to neuroepithelial cells

CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a cDNA encoding a  
 CC novel human polypeptide of the invention.  
 XX SQ Sequence 906 BP; 220 A; 225 C; 216 G; 245 T; 0 other;  
 Query Match 20.2%; Score 699.2; DB 22; Length 906;  
 Best Local Similarity 98.9%; Pred. No. 4.6e-129;  
 Matches 704; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 OY 515 ATGGCTGAGTGAAGAACTGTAGTGTATATATCTCGAAGCAAGGCTCCGAGTTCC 574  
 DB 906 ATGGCTGTAGTGAAGAACTGTAGTGTATATATCTCGAAGCAAGGCTCCGAGTTCC 847  
 OY 575 AACTCTTCCCTAGAGTGTAGAGTGAATGCGGGGTCAAGTTATTTACTCGCAT 634  
 DB 846 AACTCTTCCCTAGAGTGTAGAGTGAATGCGGGGTCAAGTTATTTACTCGCAT 787  
 OY 635 TCACATTGATTAAGCATCTCTTCATTCCTCTCTGTGAAATGTTTCCCAAGAGAC 694  
 DB 786 TCACATTGATTAAGCATCTCTTCATTCCTCTGTGAAATGTTTCCCAAGAGAC 727  
 OY 695 ACGGCTAATGATCTACCAAGGACTCCAAAGGAAAGTTTTCATTGACAGATGATTC 754  
 DB 726 ACGGCTAATGATCTACCAAGGACTCCAAAGGAAAGTTTTCATTGACAGATGATTC 667  
 OY 755 TTGTTCCGTTATATCTGAGTATCTCAGGAGACAGGAGTGTCTCTGATCACTTT 814  
 DB 666 TTGTTCCGTTATATCTGAGTATCTCAGGAGACAGGAGTGTCTCTGATCACTTT 607  
 OY 815 CCAGAAAGAAAGAACTGAAGAAAGGAAAGCTGAATATCTCCAGCTCCCAAGTGTCAAA 874  
 DB 606 CCAGAAAGAAAGAACTGAAGAAAGGAAAGCTGAATATCTCCAGCTCCCAAGTGTCAAA 547  
 OY 875 CTCTGAGACCCCGAATGAATCAAGAAAGCCCAAGTAATTCGCAAGAGATTGAA 934  
 DB 546 CTCTGAGACCCCGAATGAATCAAGAAAGCCCAAGTAATTCGCAAGAGATTGAA 487  
 OY 935 GATGCTTCCCAAGAAAGCAACAAAGATCTGCCCCCTCTCTCTGCTCCCTCCGAC 994  
 DB 486 GATGCTTCCCAAGAAAGCAACAAAGATCTGCCCCCTCTCTCTGCTCCCTCCGAC 427  
 OY 995 GCGAAGTGGGTTTCATTAATCTGTGGTTACAGAGATCTGCACTTGGGCAAGAGGAA 1054  
 DB 426 GCGAAGTGGGTTTCATTAATCTGTGGTTACAGAGATCTGCACTTGGGCAAGAGGAA 357  
 OY 1055 CAGGAGATGCAAGATTTGCGAGAGTTCCCGGATTTTGTGTGGAAGATTTCTTGG 1114  
 DB 366 CAGGAGATGCAAGATTTGCGAGAGTTCCCGGATTTTGTGTGGAAGATTTCTTGG 307  
 OY 1115 GCAAAAGAAAGCTTTGAGAACTTTGAATGAAGACAGAGACCTTGATGAGCCCAAGAA 1174  
 DB 306 GCAAAAGAAAGCTTTGAGAACTTTGAATGAAGACAGAGACCTTGATGAGCCCAAGAA 247  
 OY 1175 AGATACCTCCCAAGTTTATCTCAATTCAAGACCTGGAAGAGGCTTTTG 1226  
 DB 246 AGATACCTCCCAAGTTTATCTCAATTCAAGACCTGGAAGAGGCTTTTG 195  
 RESULT 5  
 AAS34230  
 ID AAS34230 standard; cDNA; 440 BP.  
 XX AAS34230;  
 XX 17-DEC-2001 (first entry)  
 XX Human cDNA encoding a novel foetal antigen, SEQ ID No 754.  
 XX Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;  
 XX immunomodulator; cardiovascular; cytoskeletal; nephrothropic;  
 KW

KW Cardiovascular; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; breast neoplasm; cancer;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.  
OS Homo sapiens.  
PN WO200155312-A2.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01321.  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 19-APR-2000; 2000US-0198122.  
XX 19-MAY-2000; 2000US-0205215.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216447.  
XX 11-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225457.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226881.  
XX 22-AUG-2000; 2000US-0226886.  
XX 22-AUG-2000; 2000US-0227182.  
XX 30-AUG-2000; 2000US-0228282.  
XX 01-SEP-2000; 2000US-0228287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 05-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 06-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231245.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0231415.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0233968.  
XX 14-SEP-2000; 2000US-023397.  
XX 14-SEP-2000; 2000US-023398.  
XX 14-SEP-2000; 2000US-023399.  
XX 14-SEP-2000; 2000US-023400.  
XX 14-SEP-2000; 2000US-023401.  
XX 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239933.  
PR 13-OCT-2000; 2000US-0239934.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241285.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-024475.  
PR 08-NOV-2000; 2000US-024476.  
PR 08-NOV-2000; 2000US-024477.  
PR 08-NOV-2000; 2000US-024478.  
PR 08-NOV-2000; 2000US-024523.  
PR 08-NOV-2000; 2000US-024524.  
PR 08-NOV-2000; 2000US-024525.  
PR 08-NOV-2000; 2000US-024526.  
PR 08-NOV-2000; 2000US-024527.  
PR 08-NOV-2000; 2000US-024528.  
PR 08-NOV-2000; 2000US-024532.  
PR 08-NOV-2000; 2000US-024533.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.



PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251890.  
 PR 08-DEC-2000; 2000US-0251890.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Rosen CA, Barash SC, Ruben SM,  
 XX MPI; 2001-488782/53.  
 DR P-PSDB; AAU21410.  
 XX  
 PT New polynucleotides and polypeptides for diagnosing, treating,  
 PT preventing or prognosing e.g. diseases or disorders of the nervous,  
 PT musculoskeletal, excretory, gastrointestinal, reproductive, and  
 PT respiratory systems -  
 Claim 1, SEQ ID No 754; 642bp; English.  
 XX  
 CC The invention relates to novel nucleic acids encoding novel human foetal  
 CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
 CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. The antibodies to the antigens can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC include autoimmune diseases e.g. rheumatoid arthritis, or  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver.  
 CC cardiovascular diseases e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiodysgenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. Numerous  
 CC examples of diseases and disorders treated by the nucleic acids  
 CC and proteins are given in the specification. The present sequence

Query Match 12.28; Score 423.4; DB 22; Length 440;  
 Best Local Similarity 98.94; Pred No. 1.7e-74;  
 Matches 435; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1842 TTGGAATTAATCTCAATTAAGGAAATTTAAAGGAAATTAATCAATATAT 1901  
 DB 1 TTGGAATTAATCTCTTAAGGAAATTTAAAGGAAATTAATCAATATAT 60  
 QY 1902 GCAATTTCTTGAACCAATATGATATATATATATATATATATATATAT 1961  
 DB 61 GCAATTTCTTGAACCAATATGATATATATATATATATATATATATATAT 120  
 QY 1962 CTAAACATGTAATTCACAGGGTAGATTTCTTCAATGTAAGTACAGAAATCTT 2021  
 DB 121 CTAAACATGTAATTCACAGGGTAGATTTCTTCAATGTAAGTACAGAAATCTT 180  
 QY 2022 TTTAGTATTTGTTTCTTACTTCTCCATGTCATATATATATATATATATAT 2081  
 DB 181 TTTAGTATTTGTTTCTTACTTCTCCATGTCATATATATATATATATATAT 240  
 QY 2082 CCAGTACGTAAGTAAGTACAGAGGCTTGGAGATCTTATATCCAACTGGG-TTTT 2140  
 DB 241 CCAGTACGTAAGTAAGTACAGAGGCTTGGAGATCTTATATCCAACTGGGTTTTT 300  
 QY 2141 TCTTCATCTTCTACCTCCCTCTTGTGATAGGATGTAAGTAAGTAAGTATGGCCCA 2200  
 DB 301 TCTTCATCTTCTACCTCCCTCTTGTGATAGGATGTAAGTAAGTAAGTATGGCCCA 360

QY 2201 ATGCAATTAATTTGAAATTTTATTTGTTTCTTCTTCTTATATAGGGGTTGGGGGA 2260  
 DB 361 ATGCAATTAATTTGAAATTTTATTTGTTTCTTCTTCTTATATAGGGGTTGGGGGA 420  
 QY 2261 ATGCAATTAATTTATATATATATATATATATATATATATATATATATATAT 2280  
 DB 421 ATGCAATTAATTTATATATATATATATATATATATATATATATATATATAT 440

RESULT 6  
 AAD49513  
 ID AAD49513 standard; cDNA; 2398 BP.  
 XX  
 AC AAD49513;  
 XX  
 DT 24-MAR-2003 (first entry)  
 XX  
 DE Human TRICH-15 cDNA.  
 XX  
 KM Human; transporter and ion channel; TRICH; atherosclerosis; cancer;  
 XX gene therapy; gene; seq.  
 XX  
 OS Homo sapiens.  
 XX  
 FH KEY Location/Qualifiers  
 FT CDS 114..1535  
 FT /\*tag= a  
 FT /product= "Human TRICH protein"  
 FT sig\_peptide 114..230  
 FT /\*tag= b  
 FT mac\_peptide 231..1532  
 FT /\*tag= c  
 FT /product= "Mature human TRICH protein"  
 XX  
 WO200283712-A2.  
 XX  
 24-OCT-2002.  
 XX  
 12-APR-2002; 2002WO-US11760.  
 XX  
 12-APR-2001; 2001US-283440P.  
 XX  
 20-APR-2001; 2001US-285592P.  
 XX  
 27-APR-2001; 2001US-287263P.  
 XX  
 04-MAY-2001; 2001US-288666P.  
 XX  
 18-MAY-2001; 2001US-292042P.  
 XX  
 25-MAY-2001; 2001US-293724P.  
 XX  
 22-JAN-2002; 2002US-351107P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA  
 BAUGH MR, ELLIOTT VS, HAFALIA AJA, YANG J, WALIA NK, RAMKUMAR J;  
 PI FORSYTHE TJ, LU Y, TANG YT, YUE H, RAUNAN BE, LAI PG, AZIMZAI Y;  
 PI LU DM, GANDHI AR, THORNTON M, NGUYEN DB, ARIZU CS, EMERLING BM;  
 PI SARNIAKAR A, YAO MG, DING L, HE A, GRIFFIN JA, SANJANWA MM;  
 PI GRETZEN KD, LEE EA, XU Y, AU-YOUNG JK, DAS D, LEE SY, CHANG H;  
 XX MPI; 2003-092996/08.  
 DR P-PSDB; AAE32081.  
 XX  
 PT New human functional transporters and ion channels (TRICH)  
 PT polypeptides, useful for preparing a composition for diagnosing or  
 PT treating a disease associated with decreased expression or  
 PT overexpression of TRICH e.g. cancer -  
 Claim 5; Page 200-201; 204pp; English.  
 CC The invention relates to human transporters and ion channels (TRICH)  
 CC polypeptides and nucleic acid molecules encoding such polypeptides.  
 CC TRICH proteins are useful for preparing compositions for diagnosing or  
 CC treating diseases or conditions associated with decreased expression  
 CC or overexpression of functional TRICH e.g. atherosclerosis or cancer.  
 CC The invention is useful in gene therapy. The present sequence is  
 CC human TRICH cDNA.



XX	Sequence 2398 BP, 644 A; 588 C; 604 G; 562 T; 0 other;	
SQ	Query Match 9.24; Score 319.4; DB 25; Length 2398; Best Local Similarity 58.04; Pred. No. 9.9e-54; Matches 769; Conservative 0; Mismatches 466; Indels 90; Gaps 8;	
OY	560 GGGTCGCGAGTTCCTCCCACTCTCCCTGAGGGGTGAGAGTGGTGAATGCGGGGCTCAATT	619
DB	216 GGGCCCTGGCGACCTCGCCCTTCCCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT	275
OY	620 TATTTTACTGCGCTTCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	679
DB	276 TATGTCACCAAGCACTGACGCTGCTGACGCTGCTGACGCTGCTGACGCTGCTGACGCTGCTGAC	335
OY	680 TCCCAAGAGAGACACGCTAAT-----GATTAACCAAGATCTCAAG	724
DB	336 TCGCCCTGATCCCG	395
OY	725 GGAAAGTTTTTCAATGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	784
DB	396 GCGCGCTTCTTCAATGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	455
OY	785 GACAGCAGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	844
DB	456 GACAGCAACTCG	515
OY	845 GAATACCTTCCAGCTCCCGACATTTGGTCAAACTCTGACCCCGACGATGAATCAAGCAAGC	904
DB	516 GAGTATTTCCAGCTCCCGACATTTGGTCAAACTCTGACCCCGACGATGAATCAAGCAAGC	575
OY	905 CC-----AGATGAATTCGCCACAGTGAATTGA-----AGATGC	939
DB	576 TTTCTCAACGACGAGGCTGCGACGAGCTGCGAGCAACCTCTGCGAGGATGACG	635
OY	940 CTCCCAAGAAAGCAACAAGATCTGCCCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	986
DB	636 GAGCGCTGCTGCTGCG	695
OY	987 -----CTGCCGACGCAAGTGGGTTTCAATTAATCTGAGTTACAGA	1027
DB	696 GGTGCG	755
OY	1028 GGATCTGCACTTTGGGCAAGAGGGAAGGCAAGCAATGCAAGTTTGGAGATTTCCCGG	1087
DB	756 GGCTCTTACACCAACCGGCGCGCAACACGAGCGACGACGACGACGACGACGACGACGACGACGACG	815
OY	1088 ATTTGGTTTGGAGAGATTTCTTGGCAAAAGATCTTGGAGAACTTTGAATGA	1147
DB	816 ATCATGTTGTCGGCGCGCATCGCGCTGCGCAAGAGTCTTCCGAGACGCTCAAGAG	875
OY	1148 AGCAGAGACCTTATGAGCGCCGAGAAAGATACCTTCAGATTTTATTCAAATTCAAG	1207
DB	876 AGCCGCGACCCGACCGGCAAGGAGATACGTCGCTTCTTCACTCAAGTTTCAAG	935
OY	1208 CACTGAAAGGCTTTTGAATATGTTGTCAAGTGTGATTTCCATGAGTGTGCTGTATAC	1267
DB	936 TACTTGAAGCAGGCTTTTATGCTGTGCGAGGCGCGCTTCCATGAGTGTGCTGTATAC	995
OY	1268 TCATGCGTGAAGCATTTTCAATCAACATATACAGATGACAAGATGCGAGCTAC	1327
DB	996 TCTTCGGGCGACCGCGCTTTCGCAACAGTACCGGACGACGACGACGACGACGACGACGACGACG	1055
OY	1328 ACTGAATATGCTTCTACGAGCT---TCCAGATGAGTACCTTCACACGAGCTGAG	1384
DB	1056 ACCGAGTACATTTTCTTCGACCACTTCAAGAAATATATATATATATATATATATATATATATATAT	1115
OY	1385 TGCTGCAAGATGCAAG---GTGACAAAGAGGAGAGGAGCGCGCTTTCGAGATGAC	1441
DB	1116 GATGAGAAACATGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1175
OY	1442 CTTCACATCTAGCTGCGACGAGTGTGAGGCTGCTCTCCAGAGACGAGCTATC	1501

DB	1176 CTCTCACTTCAGTGTGACAGACCATTCAGAGGACAGCACTCCGAGACACCATTC	1235
OY	1502 TGTGTCCCGTGACA-----GCGCAGACCAATCCAGCTCTGACCGTCCATCAAG	1555
DB	1236 AGTGCCGAGGAGCAACAGCTACCAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTAACCTTA	1295
OY	1556 AAGGCGCTGTCAGCTGATTCACAGATGACAGATGAGTGGGCGGAAAGGACCTTACTCGG	1615
DB	1296 AAGGACCTTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1355
OY	1616 ATTCTGACTTACGCTTCAGGAGATGACATGACAGCAAAAGGCTTTAAAGAA	1675
DB	1410 AAGCTTCAATTTAGAGAGAGCTGAGAAATGATTCAGGATTTCTTAATAAAAAAATT	1735
OY	1736 CCAATGCGTTTCTGAGAGAAATATCTTGGCAATCTGAATTTAAAGAAATGATCAT	1795
DB	1470 CCAATGATTTTTCAGAGGCGCAACGCAATGCGAATCTGAATGTTGAGAAATGAG	1529
OY	1796 CTATA 1800	
DB	1530 TTATA 1534	

RESULT 7  
ID ABV99059 standard; cDNA; 632 BP.

ABV99059;

14-JUN-2003 (first entry)

Human pancreatic cancer expressed cDNA SEQ ID NO 4467.

Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;

cytotoxic; tumour; gene; ss.

Homo sapiens.

WO20026317-A2.

08-AUG-2002.

30-JUN-2002; 2002NO-US02781.

30-JUN-2001; 2001US-265305P.

31-JUN-2001; 2001US-265862P.

09-FEB-2001; 2001US-267568P.

21-MAR-2001; 2001US-278651P.

28-APR-2001; 2001US-287112P.

16-MAY-2001; 2001US-291631P.

12-JUL-2001; 2001US-305484P.

20-AUG-2001; 2001US-313999P.

27-NOV-2001; 2001US-333626P.

(CORI-) CORIXA CORP.

Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;

WPI; 2002-627435/67.

New isolated polynucleotide and pancreatic tumor polypeptides, useful for diagnosing, preventing and/or treating cancer, particularly pancreatic cancer.

Claim 1: SEQ ID NO 4467; 300pp + Sequence Listing; English.

The invention relates to an isolated polynucleotide (1) comprising: (a)

any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145);

(b) complements of (a); (c) sequences consisting of at least 20

CC contiguous residues of (a); (d) sequences that hybridize to (a), under  
 CC moderately stringent conditions; (e) sequences having at least 75% or 90%  
 CC identity to (a); or (f) degenerate variants of (a). Polypeptides  
 CC (ABP6596-ABP6637) encoded by (1) and oligonucleotide can be used to  
 CC detect cancer in a patient and compositions comprising polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations and  
 CC antigen presenting cells expressing the polypeptide are useful in  
 CC treating pancreatic cancer and stimulating an immune response. The  
 CC polynucleotides can be used as probes or primers for nucleic acid  
 CC hybridization, in the design and preparation of ribozyme molecules for  
 CC inhibiting expression of the tumour polypeptides and proteins in the  
 CC tumour cells, in vaccines and for gene therapy.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 632 BP; 178 A; 92 C; 119 G; 239 T; 4 other;

Query Match 5.9%; Score 205; DB 24; Length 632;  
 Best Local Similarity 93.4%; Pred. No. 3,46-31;  
 Matches 214; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3206 AGGTACCAATAGCTCTTTCATGAGCTTGTGCTACAGAGGTTAAAGCCAGTTTATT 3265  
 DB 15 AGGTACCAATAGCTCTTTCATGAGCTTGTGCTACAGAGGTTAAAGCCAGTTTATT 74  
 QY 3266 TTGAGCATTCCTCATGATGATTTCAAGTGTAAACCAAAATTAATTTGCAATTAATAGTTGT 3325  
 DB 75 TTGAGCATTCCTCATGATGATTTCAAGTGTAAACCAAAATTAATTTGCAATTAATAGTTGT 134  
 QY 3326 GTGCCAAGACATCCCTAATTTGTTTATGCGTGTGTCATGCTGTATGCTATGATCACA 3385  
 DB 135 GTGCCAAGACATCCCTAATTTGTTTATGCGTGTGTCATGCTGTATGCTATGATCACA 194  
 QY 3386 GGTATTAAGGCAATGTGATTAATAAAAAAAAAAAAAAAAAAAAAA 3434  
 DB 195 GGTATTAAGGCAATGTGATTAATAAAAAAAAAAAAAAAAAAAAAA 243

RESULT 8  
 ABV95156  
 ID ABV95156 standard; cDNA; 614 BP.  
 AC ABV95156;  
 XX 14-JAN-2003 (first entry)

XX Human pancreatic cancer expressed cDNA SEQ ID NO 564.  
 XX Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;  
 XX cytostatic; tumour; gene; ss.  
 OS Homo sapiens.  
 XX  
 XX WO200260317-A2.  
 PN 08-AUG-2002.  
 PD 30-JAN-2002; 2002MO-US02781.  
 XX  
 XX 30-JAN-2001; 2001US-265305P.  
 PR 31-JAN-2001; 2001US-265305P.  
 PR 29-FEB-2001; 2001US-265305P.  
 PR 21-MAR-2001; 2001US-278651P.  
 PR 18-APR-2001; 2001US-287112P.  
 PR 16-MAY-2001; 2001US-291651P.  
 PR 12-JUL-2001; 2001US-305484P.  
 PR 20-AUG-2001; 2001US-313959P.  
 PR 27-NOV-2001; 2001US-333626P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;  
 PI

XX WPI: 2002-627435/67.  
 DR New isolated polynucleotide and pancreatic tumor polypeptides, useful  
 XX for diagnosing, preventing and/or treating cancer, particularly  
 PT pancreatic cancer.  
 XX  
 XX Claim 1; SEQ ID NO 564; 300pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (1) comprising: (a)  
 CC any of a group of over 4000 nucleotide sequences (ABV94628-ABV95145);  
 CC (b) complements of (a); (c) sequences consisting of at least 20  
 CC contiguous residues of (a); (d) sequences that hybridize to (a), under  
 CC moderately stringent conditions; (e) sequences having at least 75% or 90%  
 CC identity to (a); or (f) degenerate variants of (a). Polypeptides  
 CC (ABP6596-ABP6637) encoded by (1) and oligonucleotide can be used to  
 CC detect cancer in a patient and compositions comprising polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations and  
 CC antigen presenting cells expressing the polypeptide are useful in  
 CC treating pancreatic cancer and stimulating an immune response. The  
 CC polynucleotides can be used as probes or primers for nucleic acid  
 CC hybridization, in the design and preparation of ribozyme molecules for  
 CC inhibiting expression of the tumour polypeptides and proteins in the  
 CC tumour cells, in vaccines and for gene therapy.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 614 BP; 177 A; 87 C; 110 G; 236 T; 4 other;

Query Match 5.8%; Score 201; DB 24; Length 614;  
 Best Local Similarity 93.3%; Pred. No. 2,16-30;  
 Matches 210; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3210 ACCAATAGCTCTTTCATGAGCTTGTGCTACAGAGGTTAAAGCCAGTTTATTTC 3269  
 DB 1 ACCAATAGCTCTTTCATGAGCTTGTGCTACAGAGGTTAAAGCCAGTTTATTTC 60  
 QY 3270 GCAATTCCTCATGATGATTTCAAGTGTAAACCAAAATTAATTTGCAATTAATAGTTGT 3329  
 DB 61 GCAATTCCTCATGATGATTTCAAGTGTAAACCAAAATTAATTTGCAATTAATAGTTGT 120  
 QY 3330 CAAAGCATCTCTAATTTGTTTATGCGTGTGTCATGCTGTATGCTATGATCACA 3389  
 DB 121 CAAAGCATCTCTAATTTGTTTATGCGTGTGTCATGCTGTATGCTATGATCACA 180  
 QY 3390 ATAAAGCAATGTGATTAATAAAAAAAAAAAAAAAAAAAAAA 3434  
 DB 181 ATAAAGCAATGTGATTAATAAAAAAAAAAAAAAAAAAAAAA 225

RESULT 9  
 ABT09812  
 ID ABT09812 standard; cDNA; 2052 BP.  
 AC ABT09812;  
 XX 05-DEC-2002 (first entry)

XX Polynucleotide encoding the K-beta M6 protein SEQ ID NO 1.  
 XX  
 XX Cytostatic; cardiact; neuroprotective; immunomodulator; antimigraine;  
 XX sedative; gynaecological; potassium channel beta subunit; K-beta M6;  
 XX gastrointestinal; reproductive; neural; sleep; low DNA repair capacity;  
 XX hyperproliferation channel activity; cardiovascular; melatonin synthesis;  
 XX mammary cancer tumorigenesis; pineal gland associated disorder;  
 XX pulmonary disorder; immune disorder; NF-kB activity; migraine headache;  
 XX low free-radical buffering capacity; delayed sleep phase syndrome;  
 XX circadian cycle; melatonin secretion; cancer; gene; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200270727-A2.  
 PN

12-SEP-2002.

21-FEB-2002; 2002WO-US05674.

21-FEB-2001; 2001US-270132P.

27-MAR-2001; 2001US-278953P.

(BRIM ) BRISTOL-MYERS SQUIBB CO.

Feder J, Lee L, Chen J, Jackson DG, Ramanathan C, Siemers N, Chang H;

WPI; 2002-713455/77.

P-PDB; AB010886.

New polynucleotide encoding human potassium channel beta subunit polypeptide, useful for diagnosing, preventing, treating or ameliorating e.g. cancer

Claim 1; Fig 1; 332pp; English.

The invention relates to an isolated polynucleotide encoding a potassium channel beta subunit (K+beta6) polypeptide or its variants. The human potassium beta subunit polynucleotide or polypeptide is useful for diagnosing, preventing, treating or ameliorating a pathological condition such as gastrointestinal, reproductive, neural, sleep, cardiovascular or pulmonary disorders, a disorder related to hyperpotassium channel activity, an immune disorder related to aberrant NF-kB activity, pineal gland associated disorders, migraine headaches, disorders associated with aberrant melatonin synthesis and/or release or with low DNA repair capacities or low free-radical buffering capacity, delayed sleep phase syndrome, aberrations in circadian cycle, mammary cancer tumorigenesis, age related disorders associated with decreased melatonin secretion, or cancer. This polynucleotide sequence represents the cDNA encoding the potassium channel beta subunit (K+beta6) protein of the invention.

Sequence 2052 BP; 380 A; 640 C; 607 G; 425 T; 0 other;

Query Match 4.8%; Score 157; DB 24; Length 2052;  
Best Local Similarity 64.6%; Pred. No. 1.5e-23;  
Matches 267; Conservative 0; Mismatches 140; Indels 6; Gaps 1

967 CCCCCCTTCTCCTCCTCCTCCCTCCGACCGCAGTGGGGTTTCATTGCTGAGTTACG 1020  
705 CACGCGCGTCCACAGTCCCTGAGCGGAGCGCGCGCTCGGCTACATCACCATTCCGCTACCG 764  
1027 AGGATCTCTGCACCTTGGGCGAGAGGAGGACAGGCGAGATTCGGAAGTTCCGAGAGTTCGCCG 1081  
765 CGGCTCTTACACATCTGGCGGCGGAGCGCGCAGGCGGAGCCCAAGTTCCGGCAGATGCGCGC 824  
1087 GATTTTGGTTTGTGGAAGATTTCTTGGGAAAAGATTTTGTGGAGAACTTGGATGCA 1144  
825 CATTCACGTTTGGCGAAGAACGTCGCTCGCCMAAGATGTTTGGGAGAACCTTGACGA 884  
1147 AAGCAGACACCTTGATGAGGACCCCGAAGAAAGATACACTCCAGATTTTATCTCAATCAATCA 1200  
885 AAGCGCGGAGCCCGACGTCGCTCCCGGACGCTACACTCCGCGTATTACTCTCAAGTTCAA 944  
1207 GCACTGTGAAAGGGCTTTTGATATATTTCTTCAGATGATGCAATTCAGATGGGCTCTGA 1266  
945 CTTCCTGAGACAGGCTTTCGACAAAGTGTCCGATGCGGAGCTTTCACATATGGGAGTCCAG 1000  
1267 CTGATCGGTGACGACATCTTT-----CATTAACCAATATATACAGATGAACAAGATCTGTC 1322  
1005 CTCACCGGAGCACTGGCGCTTTGCCAGACGACCCGACAGAGGAGGAGCAAGAGATCTGAC 1066  
1321 AAGCTACACTGAATATGTTCTTCTTACCGGTGAGCTTCCAGATGGTCAACCTTCAG 1373  
1065 CAGCTACACCGAGTACGTTCTTCTTCGAGGAGATGAGCTCCCAAGACCCCTCTGCG 1117

ID	AB08125/c
AC	AB08125 standard; cDNA; 109201 BP.
XX	
XX	AB08125;
DT	18-SEP-2002 (first entry)
DE	Human osteoblast differentiation related CDNA SEQ ID NO 32.
XX	
KW	Human; osteoblast; stem cell differentiation; bone tissue deposition; osteoporosis; osteopathic; ss.
OS	Homo sapiens.
XX	
FN	WO2002050301-A2.
PD	27-JUN-2002.
PF	18-DEC-2001; 2001WO-US48276.
XX	
PR	18-DEC-2000; 2000US-255882P.
XX	24-APR-2001; 2001US-285691P.
PA	(GENE-) GENE LOGIC INC. (PROC) PROCTER & GAMBLE CO.
P1	Jl D. Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A; Mertz L; WPI; 2002-557663/59.
DR	
PT	Use of genes and their expression profiles associated with osteoblast differentiation for screening modulators bone formation, for diagnosing or treating e.g. osteoporosis, or as markers for the differentiation process -
PS	Claim 1; SEQ ID NO 32; 78bp + Sequence Listing; English.
XX	
CC	The invention relates to genes and their expression profiles are used for: (a) screening modulators of precursor stem cell differentiation into osteoblasts, or bone tissue deposition; (b) diagnosing abnormal deposition of bone tissue, abnormal rate of osteoblast formation or osteoporosis; or (c) treating or monitoring treatment of the conditions cited in (b), or monitoring the progression of bone tissue deposition. Specific conditions include postmenopausal osteoporosis, glucocorticoid osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-induced abnormalities in bone formation or bone loss, conditions that involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis), skeletal disease linked to breast cancer, mastocytosis, Paget's syndrome or fibrous dysplasia. The present sequence is that of an osteoblast differentiation associated CDNA marker of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SO	Sequence 109201 BP; 32871 A; 23488 C; 22108 G; 30734 T; 0 other;
Query Match	4.8%; Score 167; DB 24; Length 109201;
Best Local Similarity	64.8%; Pred. No. 3,5e-23;
Matches 267;	Conservative 0; Mismatches 140; Indels 6; Gaps 1
Dy	967 CCCCCCTCCTCCCTGCTCCCTCGGCAGACGGCAAGTGGGGTTTCATTAAGTGCGTTACG 10286
Db	9291 CAGCCGCTCCCAACTGCTGTGAAGCGCACCGCGGCGCTCGGGCTAACAACATCGGCTAACG 92333
Dy	1027 AGAATCTTGACACTTGGGCAAGAGGAGCAGGCAAGATGCCAAAGTTTGGAGAAGTTCCCG 10866
Db	9231 CGGGCTCTTCAACCATGTGGCGGAGACGGGAGCGGAGCGCAAGATTCCGGGGAATGGCGG 91722
Dy	1087 GATTGTGGTTTGGAGAGATTTCCTTGGCAAAAAGAGTCTTTGGAGAAAATTGGATGA 11466

Db 9171 CATCACCGTTTGGGAAAGACGTGCTGGCCAGAGGCTTTGGGAGAACCTTAAACGA 9112  
 Qy 1147 AAGCAGAGACCTGATGAGCCCGCAGAAAGATACACCTCCAGATTTTATCTCAATTCAA 1206  
 CC 9111 AGCCGGGAGCCCGACCTGCTCCCGAGGCGTACACCTCCGGCTATTAATCTCAAGTTCAA 9052  
 Db 1207 GCACCTGGAAGGCTTTGATATGTTGTCAGAGTGGATTCACATATGATGGCTGPA 1266  
 Qy 9051 CTTCCTGAGCAGGCTTTCACAGCTGTCCAGATCGGGCTTCCATGCTGCTGAG 8992  
 Db 1267 CTGATCGGTGACAGATCTTT-----CATCAACCATATACAGATGACAAATCTGTG 1320  
 Qy 8991 CTCACAGGACCTGCGCTTTCGACAGCAGCAGCAGCAGGAGGACAGATCTGAGAC 8932  
 Qy 1321 AGCTACACGTAAATATGTCTTCTACCGTGAGCTTCCAGATGTCAACCTCAAC 1373  
 8931 CAGCTACACGAGTACGTCTTCTGACAGGAGTGAAGCTCCCGACACCCCTCCG 8879  
 RESULT 11  
 ID ABQ40654  
 ID ABQ40654 standard; DNA; 854 BP.  
 AC ABQ40654;  
 DT 12-JUL-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27245.  
 XX  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 OS Homo sapiens.  
 XX  
 XX WO200218632-A2.  
 PN 07-MAR-2002.  
 XX  
 XX 01-SEP-2001; 2001WO-EP10074.  
 PF  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 ) (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX WPI; 2002-371829/40.  
 DR  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation of side effects of  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX  
 SO Sequence 854 BP; 131 A; 98 C; 289 G; 336 T; 0. other;  
 Query Match 3.3%; Score 114.2; DB 24; Length 854;  
 Best Local Similarity 59.7%; Pred. No. 3.6e-13;  
 Matches 213; Conservative 0; Mismatches 138; Indels 6; Gaps 1;  
 Qy 1001 TGGGTTTCTACTGCTGCTGATCAGAGATCTCGACCTTGGGACAGAGGACAGCA 1060  
 Db 491 TCGGTTATATTAATTAATGCTTATCGGCTTTATATATATTCGGGGGACCGTACGG 550  
 Qy 1061 GATGCCAAGTTTCGAGAGATTCGCCGAGATTTGGTTGTGGAAGATTTCTTGGCAAA 1120  
 Db 551 GACGTTAAGTTTCGCGAGAGTGGCGGTATTAATTCGTTGCGAAGACGTGTTGTTAG 610  
 Qy 1121 GAAGTCTTTGAGAACTTTGAAATGAAGACAGACCCCTATGACCCCGAAGAAATAC 1180  
 Db 611 GAGGTGTTGGGATATTTTGAACGAAGTCGGATTTTCATGCTTTTCGAGAGCTTAT 670  
 Qy 1181 ACCTCAGATTTTATCTCAATTCAGACACCTGGAAGGCTTTGATATGTTGTCAAG 1240  
 Db 671 ATTTGCGTTATTAATTTTAAGTTTAATTTTGGAGTGGTTTCGATAAGTTGTTCAG 730  
 Qy 1241 TGTGATTCACATGTTGGCTGTATCATGCTGATGAGATCTTT-----CATCAAC 1294  
 Db 731 TCGGTTTATTAATGATGCGGTGCTGATTTTACGGGTATTTGCGTTTGTAGTATAC 790  
 Qy 1295 CAAATACAGATGACAGATCTGCTCAAGCTCACTGAATATGCTTCAACGTGAG 1351  
 Db 791 GATTAGACGAGATTAAGATTTGATTAATGATTAATACGATGACGTTTTTGTAGGAG 847  
 RESULT 12  
 ID ABQ40655/C  
 ID ABQ40655 standard; DNA; 854 BP.  
 AC ABQ40655;  
 DT 12-JUL-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27246.  
 XX  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 OS Homo sapiens.  
 XX  
 XX WO200218632-A2.  
 PN 07-MAR-2002.  
 XX  
 XX 01-SEP-2001; 2001WO-EP10074.  
 PF  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 ) (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX WPI; 2002-371829/40.  
 DR  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.

XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC A011410-A0605121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

Sequence 854 BP; 336 A; 289 C; 98 G; 131 T; 0 other;

Query Match	Score	DB	Length
3.38;	114.2;	24;	854;

Matches 213; Conservative 0; Mismatches 138; Indels 6; Gaps 1

QY	1001	TGGGGTTTCAATTACTGTGGGTACAGAGAAATCTGGACCTTGCGGACAGAGAGGACAGGCA	1060
Db	364	TGGGGTTATTATTATTCGGTTATGCGCGTTTTTATATTAATGAGGCGGAGACGGTGGGCG	305
QY	1061	GATGCGAAGTTTGGAGATTTCCCGGATTTTGGTTTGTGMAAGATTTTCTTGGGCAAA	1120
Db	304	GACGTTAAGTTTGGGGAGTGGCGGTTATATCGTTTGGCGAAAGAACGTGTTGGTTAAG	245
QY	1121	GAAAGCTTTSAGAAACTTGAATGAAGACAGAGACCCTGATCGAGCGCCGAGAAATAC	1180
Db	244	GAGGTGTTGGGAAATTTTGAACGAAAGTCGGGATTTTGATCGTTTTTGGAGCGTTAT	185
QY	1181	ACCCGACGATTTTATCTCAATTCAGACCTGCGAAAGGCTTTGATATGTTCTCAGAG	1240
Db	184	ATTTCGCGTTATTATTATTAAGTTTAATTTTTTGGAGTAGAGTTTTCGATTAAGTTGTTCGAG	125
QY	1241	TCTGGAATTCACATGCTGGCCCTGTAACTCATCGGTACAGCATCTTT-----CATCAAC	1294
Db	124	TGGGGTTTTTAAATGATGGCGGTGAAGTTTACGAGATTTTGCATTTTGTAGTAAATAC	65
QY	1295	CATATTAAGATGACAAAGTCTGTCAAGCTACATGAAATATGTCTCTACCGGTAG	1351
	64	GATTTAGACGAGATTAAGATTTGGATTAGTTATATGAGATGACGTTTTTTTGTAGAGAG	8

### RESULT 13

ID ABQ13668 standard; DNA; 1757 BP.

AC	ABQ13668;
XX	
DT	12-JUL-2002 (first entry)

**Oligonucleotide for detecting cytosine methylation SEQ ID NO 259.**

KM Human; cytosine methylation; 5'-cpg-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism  
KM SNP; cell differentiation; ds.

**Homo sapiens**

PN WO200218632-A2.

PD 07-MAR-2002.

XX 01-SEP-2001; 2001MO-EP10074.  
PF 01-SEP-2000; 2000DE-1043826.  
XX 05-SEP-2000; 2000DE-1044543.  
PR  
PR  
XX  
XX  
PA (EPiG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40

PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomer, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC A011410-A06051121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

**SQ** Sequence 1757 BP; 246 A; 209 C; 640 G; 662 T; 0 other;

Query Match	3.2%	Score 109.8;	DB 24;	Length 1757;
-------------	------	--------------	--------	--------------

Matches 171; Conservative 0; Mismatches 102; Indels 0; Gaps 0

Qy	1001	GGGGGTTTCATTACGTGGGGTTACAGAGNATCTGCACCTTGGGACAGAGGACAGGGA	1066
Db	1498	TGGGGTTATATATATATCGTTATCGGGGTTTTATATATATCGGGCGGAGCGGTAGGGC	1544
Qy	1061	GATCGCAAGTTTCGAGAGTTCCCGGATTTGGTTGTGGAAAGATTTCTCTGGCAAAA	1120
Db	1545	GACGTAAAGTTTCGGGAGATGGCGCGTATATCGTTTTCGAAAGACGTGTTGTGTTAAG	1604
Qy	1121	GAACTCTTGGAAAACTTTGATGAAAGACAGACCTGTATCGAGGCCCGAAGAAATAC	1188
Db	1605	GAGGTGTTTGGGATATTTTGACGAAAGTCGGGATTTGATCGTTTTCGAGAGCTTAT	1666
Qy	1181	ACCTCCAGATTTTATCTCAAAATCAAGACCTCTGGAAAGGCTTTGATATATTTCTCAAG	1244
Db	1665	ATTTCGCGTATTAATTTTAAGTTTAAATTTTTTGGAGTAAAGTATTTTCGATTAAGTTTTCGAG	1722
Qy	1241	TGTGATTCACATGCTGGCTGTACTATCG	1273
Db	1725	TGGGATTTTATATGTTGGCGGTATGTTTACG	1757

## RESULT 14

ID ABQ13669 standard; DNA; 1757 BP

AC ABQ13669;

DT 12-JUL-2002 (first entry)

XX	DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 260.
KX	XX	Human; cytosine methylation; 5'-CpG-3'; unaccl; cytosine; diagnosis;
KM	XX	drug; side effect; cancer; central nervous system; cardiovascular;
KW	XX	gastrointestinal; respiratory system; single nucleotide polymorphism;
XX	XX	SNP; cell differentiation; de.
OS	XX	Homo sapiens.
PX	XX	WO200218632-A2.
PN	XX	
PD	XX	07-MAR-2002.
PF	XX	01-SEP-2001; 2001MO-EP10074.
PR	XX	01-SEP-2000; 2000DE-1043826.
XX	XX	05-SEP-2000; 2000DE-1044543.
XA	XX	(EPIC-) EPIGENOMICS AG.
XI	XX	
XJ	XX	Olek A, Piepenbrock C, Berlin K, Guectig D:
DR	XX	WPI; 2002-371829/40.
PT	XX	Determining the degree of cytosine methylation in genomic DNA, useful
PP	XX	for diagnosis and prognosis, comprises selective hybridization of
PS	XX	amplicons from chemically treated DNA -
XX	XX	Claim 12; 56dp + Sequence listing; 56dp; German.
CC	XX	This invention describes a novel method for determining the degree of
CC	XX	C methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC	XX	genomic sample of DNA. The sample is treated chemically to convert
CC	XX	cytosine (C) but not methylated-C to uracil, then part of the genomic
CC	XX	DNA that contains the target C is amplified to form a labeled amplicon.
CC	XX	The amplicon is hybridised to two classes, each with at least one
CC	XX	C member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC	XX	and the degree of hybridisation to both classes is determined from the
CC	XX	label on the amplicon. From the ratio of labels hybridised to the two
CC	XX	classes of oligomers, the degree of methylation is calculated. The method
CC	XX	is used: (i) for diagnosis and/or prognosis of side effects of
CC	XX	therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC	XX	of the central nervous, cardiovascular, gastrointestinal and respiratory
CC	XX	systems etc., particularly by detecting mutations or single nucleotide
CC	XX	polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC	XX	types and for investigating cell differentiation. The method allows the
CC	XX	methylation status of many C residues to be determined simultaneously.
CC	XX	ABQ1410-ABQ51121 represent genomic DNA sequences used to illustrate the
CC	XX	method for determining the degree of cytosine methylation described in
CC	XX	the disclosure of the invention.
SQ	XX	
		Sequence 1757 BP; 662 A; 640 C; 209 G; 246 T; 0 other;
		Query Match 3.2%; Score 109.8; DB 24; Length 1757;
		Best Local Similarity 62.6%; Pred. No. 3.le-12;
		Matches 171; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
OY		1001 TGGGGTTTCATTACTCGTGGGTATACAGAGAATCCCGCACTGGGCAGAGAGGACACAGCA 1060
DB		273 TCGGGTTAATAATTAATATCGATTATCGCGTTTTTAATTAATATCGGGCGAGACCGTGAGGCC 214
OY		1061 GATGCCAGTTGGAGAGATTCGCCGATTTTGGTTTGGAAAGATTTCTTGGCAAA 1120
DB		213 GACGTTTAGTTTCGGCGAGATGGCGGCTATTATCTGTTTGGCAAACAATCGTTGGTAAAG 154
OY		1121 GAAGCTTTGGAGAAACTTTGAATGAAGAAGAACCCGTATCGAGCCCCAGAAAATATC 1180
DB		153 GAGGAGTTTGGGAGATATTTGAACAAAGTGGGAGATTTCGATCGTTTTTGGAGCGTAT 94
OY		1181 ACCTCACAGATTATTTCCTCAATTCACAGCACTCGAGAAAGGACTTTTGAATATGTTGCAGAG 1240
DB		93 ATTTCGCGATTATATTTTAAAGTTTAAATTTTGGAGATAGATTTTGGAAATGATTTTCGAG 34

1241 GTGATTCACATGTCGACCTGTAATCATCTC 1273  
Db 33 TCGGTTTATATGTCGCGCGTAGTTTACG 1

RESULT 15  
ID ABO40656/C  
XX ABO40656 standard; DNA; 854 BP.  
XX ABO40656;  
XX 12-JUL-2002 (first entry)  
XX Oligonucleotide for detecting cytosine methylation SSO ID NO 27247.  
DE  
RW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
RW drug; side effect; cancer; central nervous system; cardiovascular;  
RW gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX Homo sapiens.  
OS  
XX WO200218632-A2.  
XX  
XX 07-MAR-2002.  
XX  
XX 01-SEP-2001; 2001NO-EP10074.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX 03-SEP-2000; 2000DE-1044543.  
XX  
XX (EPig-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K, Gnetig D;  
PI WPI; 2002-371829/40.  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful  
XX for diagnosis and prognosis, comprises selective hybridization of  
XX amplicons from chemically treated DNA -  
XX  
XX Claim 12; 56bp + Sequence listing; 56bp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridization to both classes is determined from the label on the amplicon. From the ratio of labels hybridized to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABO1310-ABO5412 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 854 BP; 132 A; 98 C; 302 G; 322 T; 0 other;

Query Match 3.1%; Score 108.6; DB 24; Length 854;  
Bee Local Similarity 57.1%; Pred. No. 4.6e-12;  
Matches 220; Conservative 0; Mismatches 159; Indels 6; Gaps 1;

967 CCGCCCTTCCTCCCTGCTCCCTCCGACGACCGAAGTGGGGTTTCATTACTGTGGATTACG 1026

Db 338 CACGCCGTCCTAACGACACCGAGCTCGAATCATCACCATCGACTACG 339  
Qy 1027 AGATCCTGACCTTGGGAGAGGAGAGGAGATGCCAAGTTGAGAGTCCCG 1086  
Db 338 CGACTCCTACACCATCGAACGAAAGCGCAACGAAAGCCAAATTCGACGATACGCG 279  
Qy 1087 GATTTGGTTTGGAGAGATTCTTGGCAAAAGCTTTGGAGAACTTTGAATGA 1146  
Db 278 CATCACCGTTTAGAABAAAAGCTCCCTAACCAAAAATTTAAABAAACCCCTAAAG 219  
Qy 1147 AAGCAGAGACCTGATCGAGCGCCGAAAGATACCTCAGATTTATCTCAATTCA 1206  
Db 218 AAACGAAACCCGAGCGTCCCGAAGCTACACCTGCGCTATTACTGAAATTCGA 159  
Qy 1207 GCACCTGGAAGGGCTTTGATATGTTGTCAAGTGTGATTCACATGTGGCTGTAA 1266  
Db 158 CTTCCTAAACAAACCTTGCACAACTATCCGAATCGAACTTCACATATTAACGTACAA 99  
Qy 1267 CTGATGGTGACAGCATCTTT-----CATCAACCAATATACAGATGAAGAATCTGTC 1320  
Db 98 CTCACGAAACACCTACGCTTACCAACACACGACCAAAAGAAACAAATCTAAC 39  
Qy 1321 AAGCTACACTGAATATGCTTCTAC 1345  
Db 38 CAAGTACACCGAATAGCTCTTAC 14

Search completed: September 5, 2003, 15:10:11  
Job time : 850 secs